Optimale Eigenschaften von Rangtests für das Zwei-Stichproben-Problem und ihre relative Effizienz zu parametrischen Testverfahren

Inaugural-Dissertation
zur Erlangung des Grades eines
Doktors der Wirtschaftswissenschaften
beim Fachbereich Wirtschaftswissenschaft
der Freien Universität Berlin

vorgelegt von
Herbert Büning
aus Oberhausen

1973

Optimale Eigenschaften von Rangtests für das Zwei-Stichproben-Problem und ihre relative Effizienz zu parametrischen Testverfahren

Inaugural-Dissertation
zur Erlangung des Grades eines
Doktors der Wirtschaftswissenschaften
beim Fachbereich Wirtschaftswissenschaft
der Freien Universität Berlin

vorgelegt von
Herbert Büning
aus Oberhausen

Vorsitzender des Fachbereichs: Prof. Dr. G. Dlugos

Referent: Prof. Dr. H. Münzner

Korreferent: Prof. Dr. W. Wetzel

Tag der mündlichen Prüfung: 17. Dezember 1973

UNIV.
LEIDEN
BIBL.

Druck: Der Präsident der Freien Universität Berlin – Zentrale Universitätsdruckerei V aA 4, 1 Berlin 41

Inhaltsverzeichnis

		Seite
Einleitung		5
1. Kapitel	Grundlegende Begriffe der Testtheorie	
1.1	Test und Teststatistik	9
1.2	Optimale Eigenschaften	12
2. Kapitel	Das Invarianz-Prinzip	
2.1	Induziertes W-Maß und invarianter Test	18
2.2	Invariante und fundamental invariante Abbildungen und ihre Anwendung auf Testprobleme	22
2.3	Eine spezielle Abbildungsgruppe	25
3. Kapitel	Rangtests als Spezialfall nichtparametrischer Tests	
3.1	Der Begriff "nichtparametrischer Test"	32
3.2	Das Zwei-Stichproben-Problem mit einigen speziellen Alternativen	33
3.3	Rangtests - Permutationstests	38
3.4	Anwendung des Invarianz-Prinzips auf Rangtests für das einseitige Zwei-Stichproben-Problem	39
3.4.1	Herleitung einer suffizienten Rangstatistik	39
3.4.2	Verteilung der Ränge unter $F_1 = F_2$ und $F_1 \neq F_2$	42
3.5	Ein bester Rangtest für $F_2 = F_1^k$	46

Seite

4. Kapitel Lokal beste Rangtests

4.1 Definition und Nachweis der Existenz lokal bester Rangtests für parametrische Alternativen 50

4.2 Lokal beste Rangtests auf Lage und Variabilität, Beispiele 57

4.3 Lokal beste Rangtests für eine Mischungsalternative 63

4.4 Gütevergleich des besten Rangtests mit dem C_1-Test und dem U-Test bei Normalalternativen 65

5. Kapitel Relative Effizienz ausgesuchter Rangtests für das Zwei-Stichproben-Problem

5.1 Definitionen der relativen Effizienz 68

5.2 Die asymptotisch relative Effizienz unter besonderer Berücksichtigung des Pitmanschen Konzepts 72

5.3 Die Pitmansche A.R.E. einiger Rangtests zum t-Test bzw. F-Test 78

5.4 Finite relative Effizienzen des U-Tests und C_1-Tests zum t-Test für Normalalternativen 80

5.4.1 F.R.E. für m=n und einseitige Alternativen 80

5.4.2 F.R.E. für m≠n und einseitige Alternativen 85

5.4.3 F.R.E. für zweiseitige Alternativen 87

5.4.4 F.R.E. über normale Interpolation 89

5.4.5 F.R.E. des U-Tests für große m=n 91

5.4.6 Abschließende Bemerkungen 96

Literaturverzeichnis 99

Einleitung

Das Konzept der nichtparametrischen Testverfahren in der Mathematischen Statistik ist vorwiegend in jüngster Zeit erstellt und entwickelt worden, jedoch reichen einige Ansätze weit zurück. So stellt Bradley fest:[1)]

"Distribution-free statistical tests can be traced at least as far back as 1710, when John Arbuthnott used the statistical Sign test in an attempt to prove the wisdom of Divine Providence."

Zahlreiche neuere nichtparametrische Testverfahren erobern sich mehr und mehr neben den klassischen Methoden einen festen Platz in den verschiedensten praktischen Anwendungsbereichen, insbesondere dort, wo über die Verteilung der Grundgesamtheiten keinerlei Information vorliegt und somit die "traditionelle" Annahme einer Normalverteilung nicht gerechtfertigt ist.

Die Literatur zu nichtparametrischen Testverfahren ist kaum noch überschaubar. Zu denjenigen Lehrbüchern, die den nichtparametrischen Testverfahren den größten Platz einräumen oder ihnen ausschließlich gewidmet sind, zählen in erster Linie die von Hajek, Witting/Noelle, Fraser, Lehmann und Schmetterer und für die praktische Anwendung die von Lienert und Siegel.

Unter den verteilungsfreien Verfahren stellen die sogenannten Rangtests eine besonders wichtige Gruppe dar. Die Verwendung von Rängen statt der beobachteten Stichprobenwerte ist in vielen Fällen deshalb naheliegend, weil die zugrunde liegende Meßskala nur eindeutig bis auf beliebige stetige und monotone Transformationen ist (ordinale Skala), d.h. selbst nicht mehr als die Rangordnung zum Ausdruck bringt. Das liegt stets dann vor, wenn die Untersuchungsergebnisse aus Angaben von Qualitäten

1) Bradley, [9], S. 1

bestehen (Schulnoten u.a.). In den Fällen, in denen die Meßskala eindeutig bis auf lineare Transformationen ist (metrische Skala), tritt natürlich durch die Verwendung von Rangzahlen ein Informationsverlust ein. Dieser fällt jedoch nicht so stark ins Gewicht, wie es zunächst den Anschein haben könnte, auch dann nicht, wenn die Meßwerte normalverteilt sind. Diese hohe Effizienz der Rangtests zu den entsprechenden parametrischen Verfahren, wie sie im 5. Kapitel herausgestellt wird, spricht für ihre bevorzugte Anwendung, zumal dann, wenn kleine Stichprobenumfänge vorliegen.

Die Untersuchung von Rangtests, die im wesentlichen auf kombinatorischen Überlegungen beruht, ist im Vergleich mit parametrischen Verfahren auch dadurch vereinfacht, daß der betrachtete Stichprobenraum endlich ist. Darüber hinaus haben Rangtests den Vorteil, daß ihre Anwendung keine umfangreichen Berechnungen erfordert.

Die Erstellung einer Rangtest-Theorie basiert auf zwei Voraussetzungen:

1. Die Stichprobenvariablen sind unabhängig.

2. Die Verteilungsfunktionen der Grundgesamtheiten sind stetig, d.h. die (theoretische) Zuweisung der Ränge ist eindeutig.
Treten jedoch in der beobachteten Stichprobe gleich große Werte auf, so bieten sich für die Behandlung dieser sogenannten "ties" verschiedene Verfahren an.[1)]
Darauf wird aber in der vorliegenden Arbeit nicht eingegangen.

Unter den nichtparametrischen Testproblemen nimmt das sogenannte Zwei-Stichproben-Problem eine vorrangige Stellung ein, insbesondere im Zusammenhang mit Rangtestverfahren, die vom historischen Standpunkt aus

---

1) Bradley, [9], S. 48-54

betrachtet gerade für dieses Problem entwickelt worden sind.[1] Während die Nullhypothese im Zwei-Stichproben-Problem die Gleichheit der Verteilungen der beiden Grundgesamtheiten beinhaltet, sind für die Alternativhypothese die verschiedensten nichtparametrischen und parametrischen Ansätze diskutiert worden. Einige davon sind in Kapitel 3.2 angeführt, für die dann optimale Eigenschaften hergeleitet werden.

Im 1. Kapitel werden später verwendete Grundbegriffe der Testtheorie mit einigen Optimalitätsdefinitionen zusammengestellt.

Im 2. Kapitel wird die Reduktion der beobachteten Stichprobenwerte auf die zu ihnen gehörigen Ränge als fundamental invariante Abbildung nachgewiesen und damit die Bedeutung der Rangtests für invariante Testprobleme herausgestellt.

Im 3. Kapitel werden die aus dem 2. Kapitel gewonnenen Ergebnisse auf das einseitige Zwei-Stichproben-Problem angewandt und die Wahrscheinlichkeitsmaße als Verteilungen der Ränge unter der Null- und Alternativhypothese bestimmt, um einen besten Rangtest zu konstruieren.

Im 4. Kapitel werden lokal beste Rangtests als Verallgemeinerung gleichmäßig bester Tests für die parametrischen Alternativen aus Kapitel 3.2 hergeleitet.

Im 5. Kapitel wird ein Vergleich zwischen einigen ausgewählten Rangtests und dem t-Test bzw. F-Test angestellt, und zwar über finite und asymptotische relative Effizienzen, auch für den Fall nicht-normalverteilter Grundgesamtheiten. Die angeführten Tabellen unterstreichen die oben erwähnte hohe Effizienz dieser Rangtests zu den parametrischen

---

1) Mann/Whitney, [39], S. 50

Tests. Daß überhaupt ein Effizienzverlust bei Anwendung von Rangtestverfahren in Kauf genommen werden muß, ist nach Kendall nur natürlich. "A tool which has many purposes is not usually as efficient for any one of them as a specialized tool developed solely for a single purpose." [1]

1) Kendall, [30], S. 166

1. Kapitel Grundlegende Begriffe der Testtheorie

1.1 Test und Teststatistik

Die Annahme oder das Verwerfen einer spezifizierten Hypothese im Rahmen der Testtheorie, wie überhaupt jede statistische Entscheidung, basieren auf einem Beobachtungsbefund x, der als Realisation einer Zufallsvariablen X interpretiert wird, d.h. als Wert einer auf einem Wahrscheinlichkeitsraum $(W, \mathcal{F}, P)$ definierten meßbaren Funktion. Der Meßraum $(S, \mathcal{A})$ als Bildbereich von X wird Stichprobenraum genannt und x die Stichprobe zur Zufallsvariablen X. Ein Spezialfall eines Meßraumes, der in dieser Arbeit vorrangig betrachtet wird, ist der n-dimensionale Borelsche Meßraum $(\mathbb{R}^n, \mathcal{B}^n)$, in dem jede Stichprobe $x = (x_1, \ldots, x_n)$ zur Zufallsvariablen $X = (X_1, \ldots, X_n)$[1] als Punkt des Euklidischen Raumes gekennzeichnet werden kann.

Das Tripel $(S, \mathcal{A}, P^X)$ stellt den zum Stichprobenraum gehörigen Wahrscheinlichkeitsraum dar; dabei bedeutet P^X das durch X über dem Wertebereich $(S, \mathcal{A})$ induzierte Wahrscheinlichkeitsmaß (W - Maß) :

$$\bigwedge_{A \in \mathcal{A}} P^X(A) := P(X^{-1}(A)).$$

Für den erwähnten Spezialfall $(\mathbb{R}^n, \mathcal{B}^n, F^{(n)})$ bezeichnet $F^{(n)}$ die gemeinsame Verteilung von $X_1, \ldots, X_n$.

Bei den parametrischen Testverfahren wird der Typ des durch X induzierten W-Maßes P^X als bekannt angenommen; die zu testenden Hypothesen beziehen sich auf einen (möglicherweise mehrdimensionalen)

---

1) Statt $X = (X_1, \ldots, X_n)$ wird im folgenden kürzer $X_1, \ldots, X_n$ geschrieben und die X_i werden auch Stichprobenvariable genannt.

Parameter, der P^X vollständig festlegen soll. Es sei also $(S, \mathcal{A})$ der Stichprobenraum, Ω eine Menge von zulässigen Parametern mit mindestens 2 Elementen und $\{P_\theta^X \mid \theta \in \Omega\}$[1] eine Menge von W-Maßen über $(S, \mathcal{A})$. Weiterhin seien $\omega \subset \Omega$ mit $\omega \neq \phi, \Omega - \omega \neq \phi$ und $\alpha \in \mathbb{R}$ mit $0 \leq \alpha \leq 1$ vorgegeben.

Def. 1 : Ein Test φ ist eine $\mathcal{A}$-meßbare Abbildung

$$\varphi: S \longrightarrow [0,1] .$$

φ wird Test zum Niveau α für die Hypothese
H_o: $\theta \in \omega$ (Nullhypothese) gegen die Hypothese
H_1: $\theta \in \Omega - \omega$ (Alternativhypothese) genannt, wenn gilt:

$$\bigwedge_{\theta \in \omega} E_\theta \varphi(X) = \int_S \varphi \, dP_\theta \leq \alpha .$$

φ heißt kurz Test für das Problem $(\alpha, \omega, \Omega - \omega)$.
Je nachdem, ob ω bzw. $\Omega - \omega$ ein Element oder mehrere Elemente enthalten, heißen H_o bzw. H_1 einfach oder zusammengesetzt.

Anmerkung:
Ein Test φ läßt sich wie folgt interpretieren: Für jedes $x \in S$ gibt $\varphi(x)$ die Wahrscheinlichkeit für das Ablehnen von H_o an. Für $0 < \varphi(x) < 1$ ist vor der Entscheidung, H_o abzulehnen oder nicht, noch ein zufälliges Experiment durchzuführen, das z.B. mit Hilfe einer Tabelle über (0,1) gleichverteilten Zufallsvariablen erfolgen kann und dessen mögliche Ereignisse mit den Wahrscheinlichkeiten $\varphi(x)$ und $1 - \varphi(x)$ eintreten. Ein solcher Test heißt randomisiert, der den Spezialfall des Tests $\overline{\varphi}$ in der Gestalt einer Indikatorfunktion einschließt (nicht-randomisierter Test) :

$$\overline{\varphi}(x) := I_C(x) = \begin{cases} 1 & \text{für } x \in C \\ 0 & \text{für } x \in S - C , \end{cases}$$

wobei $C \in \mathcal{A}$ kritischer Bereich von $\overline{\varphi}$ genannt wird.

---

1) Wenn keine Verwechslung zu befürchten ist, wird im folgenden für P_θ^X kürzer P_θ geschrieben.

Die in der Theorie bevorzugte Anwendung randomisierter Tests gegenüber den nicht-randomisierten liegt darin begründet, daß die Menge aller randomisierten Tests konvex ist; das gilt nicht für die Menge der nicht-randomisierten Tests. Darüber hinaus wird α durch randomisierte Tests für einfache Hypothesen H_o und H_1 "voll ausgenutzt".

Die Konstruktion eines randomisierten Tests φ führt im Zusammenhang mit der Herleitung "optimaler" Tests für einseitige Hypothesen:

$$H_o : \theta \leq \theta_o \quad \text{gegen} \quad H_1 : \theta > \theta_o \quad (\Omega \subseteq \mathbb{R})$$

häufig zur Beschreibung von φ durch eine reellwertige Statistik.

Def. 2 : Es sei $t : S \to \mathbb{R}$ eine $\mathcal{A}$-meßbare Abbildung und $k \in \mathbb{R}$ mit $o \leq k \leq 1$.

Ist φ erklärt durch

$$\varphi(x) = \begin{cases} 1 & \text{für} \quad t(x) > c \\ k & \text{für} \quad t(x) = c \\ o & \text{für} \quad t(x) < c \end{cases},$$

dann heißt t Teststatistik für φ .

Randomisierte Entscheidungen über t werden also nur auf dem Rand $R = \{x \mid t(x) = c\}$ des Ablehnungs- bzw. Annahmebereiches getroffen. Da aber für eine stetig verteilte Teststatistik t gilt :

$$P_\theta(t(x) = c) = 0, \ \theta \in \Omega ,$$

erhält man randomisierte Tests überwiegend bei diskreten Zufallsvariablen.

Def. 3 :[1] Es seien $(S, \mathcal{A})$ und $(Q, \mathcal{D})$ zwei Meßräume, $\{P_\theta\}$ eine Menge von über $(S, \mathcal{A})$ definierten W-Maßen, und $T : S \to Q$ sei $(\mathcal{A}, \mathcal{D})$ - meßbar. Die σ-Algebra $T^{-1}(\mathcal{D})$ werde mit $\mathcal{A}_o$ bezeichnet. Dann heißt T suffizient für $\{P_\theta\}$, wenn es zu jedem $A \in \mathcal{A}$ eine über S definierte $\mathcal{A}_o$ - meßbare Abbildung $P(A|T)$ gibt, so daß gilt :

$$\bigwedge_{\theta \in \Omega} P(A|T) = P_\theta(A|T) .$$

Oft gelingt eine Reduktion des Testproblems durch eine suffiziente Statistik. Eine solche Reduktion wird in Kapitel 3.4.1 durchgeführt.

1.2 Optimale Eigenschaften

In der Regel werden mehrere Tests für das Problem $(\alpha, \omega, \Omega - \omega)$ betrachtet. Es liegt dann nahe, die "Güte" dieser Tests zu untersuchen.

Def. 4 : Es sei φ ein Test für das Problem $(\alpha, \omega, \Omega - \omega)$ und $\beta_\varphi : \Omega \to [0,1]$ mit $\beta_\varphi(\theta) = E_\theta \varphi(X)$.
Dann heißt $\beta_\varphi(\theta)$ für alle $\theta \in \Omega - \omega$ Gütefunktion von φ ; $\beta_\varphi(\theta)$ wird für festes $\theta \in \Omega - \omega$ Güte von φ genannt.

Das Konzept der Güte schließt das Problem ein, unter allen Tests für $(\alpha, \omega, \Omega - \omega)$ einen solchen mit maximaler Güte zu bestimmen. Es sei also M_α die Menge aller Tests für $(\alpha, \omega, \Omega - \omega)$.

Def. 5 : Ein Test $\varphi^* \in M_\alpha$ heißt gleichmäßig bester Test für ω gegen $\Omega - \omega$, wenn gilt :

$$\bigwedge_{\theta \in \Omega - \omega} \beta_{\varphi^*}(\theta) = \sup_{\varphi \in M_\alpha} \beta_\varphi(\theta) .$$

---

1) Schmetterer, [58], S. 246

Besteht $\Omega - \omega$ nur aus einem Element, so wird φ^* ein bester Test für das angegebene Problem genannt.

Auf der Suche nach besten oder gleichmäßig besten Tests können wir uns auf Tests beschränken, die nur von einer suffizienten Statistik abhängen. Das ist die Aussage des folgenden Satzes, dessen Beweis unmittelbar aus den Eigenschaften des bedingten Erwartungswertes folgt:

Satz 1 :[1] Es sei T suffizient für $\{P_\theta \mid \theta \in \Omega\}$ und φ ein beliebiger Test für das Problem $(\alpha, \omega, \Omega - \omega)$. Dann ist

$$\psi(T(x)) := E(\varphi(X) \mid T(x))$$

ein Test, und es gilt :

$$\bigwedge_{\theta \in \Omega} E_\theta \psi(T(x)) = E_\theta \varphi(X) .$$

In der Menge M_α werden weitere spezielle Tests ausgezeichnet:

Def. 6 : Ein Test $\varphi \in M_\alpha$ heißt unverfälscht, wenn gilt :

$$\bigwedge_{\theta \in \Omega - \omega} \beta_\varphi(\theta) \geq \alpha .$$

Die Teilmenge aller unverfälschten Tests von M_α sei mit M_α^u bezeichnet, zu der z.B. der triviale Test $\varphi(x) = \alpha$ (für alle $x \in S$) gehört. Offensichtlich ist auch jeder gleichmäßig beste Test unverfälscht. Wie noch im folgenden kurz herausgestellt wird, existiert in M_α nur für spezielle Testprobleme ein gleichmäßig bester Test. Dafür gelingt aber häufiger in M_α^u der Nachweis der Existenz eines gleichmäßig besten Tests, der dementsprechend gleichmäßig bester un-

1) Witting, [68], S. 146

verfälschter Test genannt wird.

Eine Minimalforderung, die an einen Test gestellt werden muß, ist seine "Zulässigkeit".

Def. 7 : Ein Test $\varphi \in M_\alpha$ heißt zulässig, wenn kein $\varphi_1 \in M_\alpha$ existiert, so daß

$$\bigwedge_{\theta \in \Omega-\omega} \beta_{\varphi_1}(\theta) \geq \beta_\varphi(\theta) \qquad \text{und}$$

$$\bigvee_{\theta_o \in \Omega-\omega} \beta_{\varphi_1}(\theta_o) > \beta_\varphi(\theta_o) \ .$$

Offensichtlich ist jeder gleichmäßig beste Test zulässig.
Es gilt auch :

Satz 2 : Jeder gleichmäßig beste unverfälschte Test φ^* ist zulässig.

Beweis: Angenommen, φ^* sei nicht zulässig; dann existiert $\varphi_1 \in M_\alpha$ mit

$$\bigwedge_{\theta \in \Omega-\omega} \beta_{\varphi_1}(\theta) \geq \beta_{\varphi^*}(\theta) \qquad \text{und}$$

$$\bigvee_{\theta_o \in \Omega-\omega} \beta_{\varphi_1}(\theta_o) > \beta_{\varphi^*}(\theta_o) \ .$$

Dann folgt :

$$\bigwedge_{\theta \in \Omega-\omega} \beta_{\varphi_1}(\theta) \geq \beta_{\varphi^*}(\theta) \geq \beta_\varphi(\theta) \geq \alpha$$

für alle $\varphi \in M_\alpha^u$, da φ^* gleichmäßig bester Test aus M_α^u ist; d.h. also: $\varphi_1 \in M_\alpha^u$ und weiterhin

$$\bigvee_{\theta_o \in \Omega - \omega} \beta_{\varphi_1}(\theta_o) > \beta_{\varphi^*}(\theta_o) \geq \beta_{\varphi}(\theta_o) \geq \alpha$$

im Widerspruch zur Eigenschaft von φ^*. q.e.d.

Die Frage nach der Existenz eines besten oder sogar gleichmäßig besten Tests für $(\alpha, \omega, \Omega - \omega)$ kann wie folgt beantwortet werden:

(1) Für eine einfache Nullhypothese H_o und eine einfache Alternative H_1 liefert das Lemma von Neyman-Pearson einen besten Test, das hier ohne Beweis angegeben sei :

Lemma :[1] Seien P_{θ_1} und P_{θ_2} zwei W-Maße über $(S, \mathcal{A})$, dann ist der beste Test $\overline{\varphi}$ für das Problem $(\alpha, \theta_1, \theta_2)$ gegeben durch :

$$\overline{\varphi}(x) = \begin{cases} 1 & \text{für} \quad f_{\theta_2}(x) > a f_{\theta_1}(x) \\ b & \text{für} \quad f_{\theta_2}(x) = a f_{\theta_1}(x) \\ o & \text{für} \quad f_{\theta_2}(x) < a f_{\theta_1}(x) \end{cases},$$

wobei f_{θ_1} und f_{θ_2} die Radon-Nikodym-Dichten (R.-N.-Dichten) von P_{θ_1} bzw. P_{θ_2} bezüglich eines geeigneten Maßes μ (z.B. $\mu = P_{\theta_1} + P_{\theta_2}$) sind und a und $b \in [0,1]$ bestimmt sind durch

$$E_{\theta_1} \overline{\varphi}(X) = \int_S \overline{\varphi}\, dP_{\theta_1} = \int_S \overline{\varphi} f_{\theta_1}\, d\mu$$

$$= P_{\theta_1}\{f_{\theta_2}(x) > a f_{\theta_1}(x)\} + b\, P_{\theta_1}\{f_{\theta_2}(x) = a f_{\theta_1}(x)\} = \alpha .$$

---

1) Fraser, [18], S. 73

Das vorgegebene α wird hier also "voll ausgenutzt".

Ist die Menge $\{x|\ f_{\theta_2}(x) = af_{\theta_1}(x)\}$ eine μ - Nullmenge, dann ist der beste (nicht-randomisierte) Test $\overline{\varphi}$ bis auf μ - Nullmengen aus $\mathcal{A}$ eindeutig festgelegt.

(2) Die Existenz eines gleichmäßig besten Tests für einseitige Hypothesen H_o und H_1 mit $\Omega = \mathbb{R}$ ist prinzipiell gleichbedeutend damit, daß die R.-N.-Dichten der Menge $\{P_{\mathbb{R}}\}$, die durch ein σ - endliches Maß μ dominiert sein möge, monotone Dichtequotienten besitzen.[1)]

(3) Für den allgemeinen Fall zusammengesetzter Hypothesen H_o und H_1 existiert kein gleichmäßig bester Test in M_α. In solchen Situationen wird zum einen der (nicht-randomisierte) Likelihood-Quotiententest herangezogen, der allerdings nicht auf eindeutig definierten optimalen Kriterien basiert. Zum anderen ist mit der Erklärung "optimaler" Test über die Gütefunktionen insbesondere das Konzept folgender Tests als Verallgemeinerung gleichmäßig bester Tests für zusammengesetzte Hypothesen entwickelt worden:

(a) Maximin-Tests

Als Maximin-Test wird unter allen Tests $\varphi \in M_\alpha$ der Test φ_o bezeichnet, der die minimale Güte (bezüglich H_1) maximiert, d.h. φ_o ist Lösung von

$$\inf_{\theta \in \Omega-\omega} \beta_{\varphi_o}(\theta) = \sup_{\varphi \in M_\alpha} \inf_{\theta \in \Omega-\omega} \beta_{\varphi}(\theta) .^{2)}$$

---

1) Schmetterer, [58], S. 221 und S. 227

2) Witting, [68], S. 42

(b) Strenge Tests

Es sei $\bigwedge_{\theta \in \Omega-\omega} L(\theta) := \sup_{\varphi \in M_\alpha} \beta_\varphi(\theta)$.

Dann heißt $\varphi_1 \in M_\alpha$ streng[1], wenn gilt:

$$\bigwedge_{\varphi \in M_\alpha} \sup_{\theta \in \Omega-\omega} \left[L(\theta) - \beta_{\varphi_1}(\theta)\right] \leq \sup_{\theta \in \Omega-\omega} \left[L(\theta) - \beta_\varphi(\theta)\right] .$$

Offensichtlich ist ein für $(\alpha, \omega, \Omega - \omega)$ gleichmäßig bester Test ein strenger Test.

Unter gewissen Voraussetzungen[2] ist ein Maximin-Test ein strenger Test.

(c) Lokal beste Tests

Ein lokal bester Test kann als gleichmäßig bester Test für Alternativen in einer "hinreichend kleinen" Umgebung der Nullhypothese gekennzeichnet werden. In vielen praktischen und theoretischen Problemen erweist sich eine Beschränkung der Gütefunktionen auf solche Umgebungen als nützlich, werden doch in erster Linie Alternativen H_1 betrachtet, die von H_o "nicht weit" entfernt sind.

Von diesen drei wichtigsten Verallgemeinerungen des Konzepts gleichmäßig bester Tests wird im 4. Kapitel die Theorie lokal bester Rangtests im Zusammenhang mit dem Zwei-Stichproben-Problem entwickelt.

---

1) Schmetterer, [58] , S. 277
2) Schmetterer, [58] , S. 278

2. Kapitel Das Invarianz-Prinzip

2.1 Induziertes W-Maß und invarianter Test

In zahlreichen statistischen Problemstellungen, insbesondere im Zusammenhang mit der Schätz- und Testtheorie, stellt sich die Frage nach der Invarianz des Problems bei "geeigneten" Transformationen des Beobachtungsbefunds. So wird man z.B. bei der Schätzung einer Länge θ die Unabhängigkeit der zu treffenden Entscheidung von der speziellen Wahl der Maßeinheit (wie Inch oder Zentimeter) fordern; das bedeutet Invarianz gegenüber allen Transformationen $\bar{x} = ax$, $a > 0$. Oder wird zum Testen einer Hypothese der Grundgesamtheit eine Stichprobe vom Umfang n entnommen, wobei die einzelnen n Stichprobenvariablen als unabhängig angenommen werden, dann sollte die Annahme oder Ablehnung der Hypothese unabhängig von der (belanglosen) Reihenfolge der Beobachtungen $x_1, \ldots, x_n$ sein. Das bedeutet, daß die dem Test zugrunde liegende Verteilungsannahme invariant gegenüber den Permutationen von $x_1, \ldots, x_n$ ist.

Auch die im 3. Kapitel dargestellte Rangtesttheorie für das (einseitige) Zwei-Stichproben-Problem mit der Reduktion der beobachteten Stichprobenwerte $x_1, \ldots, x_n$ auf ihre zugehörigen Ränge $r_1, \ldots, r_n$ basiert auf dem Invarianz-Prinzip, das deshalb im folgenden erläutert werden soll.

Es sei wieder $(S, \mathcal{A})$ der Stichprobenraum, G eine Gruppe $(\mathcal{A}, \mathcal{A})$-meßbarer Abbildungen g von S auf sich und $P_\Omega = \{P_\theta \mid \theta \in \Omega\}$ eine Menge von über $(S, \mathcal{A})$ definierten W-Maßen mit

$$P_{\theta_1} \neq P_{\theta_2} \quad \text{für} \quad \theta_1 \neq \theta_2 \,;\, \theta_1, \theta_2 \in \Omega \,.$$

Für jedes $g \in G$ wird durch

(1) $P_\theta^g(A) := P_\theta(g^{-1}A)$, $\theta \in \Omega$, $A \in \mathcal{A}$
eine Menge von W-Maßen $P_\Omega^g = \{ P_\theta^g \mid \theta \in \Omega \}$
über $(S, \mathcal{A})$ induziert. Wird A durch gA ersetzt, so ergibt sich aus (1) :

(1*) $P_\theta^g(gA) = P_\theta(A)$.

In der Regel gilt $P_\Omega^g \neq P_\Omega$. Das Invarianzprinzip basiert nun auf der Gleichheit der beiden Mengen von W-Maßen:

Def. 8 : P_Ω heißt invariant bezüglich einer Gruppe G , wenn gilt:

$$\bigwedge_{g \in G} P_\Omega^g = P_\Omega .$$

Bei invariantem P_Ω gibt es zu jedem $g \in G$ genau ein Element $\bar{g}\theta \in \Omega$ mit $P_\theta^g = P_{\bar{g}\theta}$;
denn sei $\bar{g}\theta_1 = \bar{g}\theta_2$ für alle $A \in \mathcal{A}$, so gilt :

$P_{\theta_1}^g(gA) = P_{\theta_2}^g(gA)$ und damit nach (1*) :

$P_{\theta_1}(A) = P_{\theta_2}(A)$, d.h. $\theta_1 = \theta_2$ wegen der vorausgesetzten Identifizierbarkeit der Parameter.

Die Abbildung $g: S \to S$ induziert also vermöge (1) unter der Voraussetzung der Invarianz von P_Ω eine Abbildung $\bar{g}: \Omega \to \Omega$, die bijektiv ist. Somit läßt sich (1) auch in der Form schreiben:

(2) $P_{\bar{g}\theta}(A) = P_\theta(g^{-1}A)$.

Bei Annahme der Invarianz von P_Ω unter $g \in G$ gilt weiterhin:

Satz 3: Die Menge $\bar{G}$ aller $\bar{g}$ bildet eine Gruppe.[1)]

Beweis: Es genügt zu zeigen, daß $\bar{G}$ homomorphes Bild von G ist, da das homomorphe Bild einer Gruppe wieder eine Gruppe ist.[2)]

Seien $g_1, g_2 \in G$ beliebig gewählt, dann gilt nach (2) :

$$P_{\overline{g_1 g_2}\,\theta}(A) = P_\theta\left((g_1 g_2)^{-1} A\right) =$$

$$P_\theta(g_2^{-1} g_1^{-1} A) = P_{\bar{g}_2 \theta}(g_1^{-1} A) = P_{\bar{g}_1 \bar{g}_2 \theta}(A)$$

für alle $\theta \in \Omega$, $A \in \mathcal{A}$, d.h. $\overline{g_1 g_2} = \bar{g}_1 \bar{g}_2$. q.e.d.

Zwei Beispiele mögen den bisher geschilderten Sachverhalt veranschaulichen:

1. Beispiel: Der Stichprobenraum sei $(\mathbb{R}, \mathcal{B})$, und es sei $P_{\mathbb{R}} := \{P_\theta \mid \theta \in \mathbb{R}\}$ die auf $(\mathbb{R}, \mathcal{B})$ definierte Menge von W-Maßen, wobei P_θ die Normalverteilung $N(\theta, 1)$ bedeutet. G bezeichne die Gruppe aller Abbildungen :

$$g : \mathbb{R} \to \mathbb{R} \qquad \text{mit}$$

$$gx = x + c, \qquad c \in \mathbb{R} .$$

Dann hat gX die Verteilung $N(\theta + c, 1)$. $P_{\mathbb{R}}$ ist invariant bezüglich G , und es gilt : $\bar{g}\theta = \theta + c$.

Im nächsten Beispiel soll gezeigt werden, daß $\bar{G}$ und G nicht notwendig isomorph sind.

---

1) Schmetterer, [58], S. 283
2) v.d.Waerden, [65], S. 38

2. Beispiel: Es sei wieder $(\mathbb{R}, \mathfrak{B})$ der Stichprobenraum, und es sei P_θ die Normalverteilung $N(0, \theta)$, $\theta \in \mathbb{R}^+$. Als Gruppe G werde gewählt: $G = \{e, g\}$, wobei g die Abbildung $g : \mathbb{R} \to \mathbb{R}$ mit $gx = -x$ bedeutet. Offensichtlich ist $P_{\mathbb{R}^+} = \{P_\theta \mid \theta \in \mathbb{R}^+\}$ invariant bezüglich G. Es gilt jedoch: $\bar{g}\theta = \theta$, d.h. $\bar{G}$ besteht nur aus dem neutralen Element $\bar{g} = \bar{e}$.

Das Konzept der Invarianz bezüglich der Testtheorie basiert neben dem Begriff der Invarianz von Wahrscheinlichkeitsmaßen auf zwei weiteren Invarianz-Begriffen:

Def. 9: Ein Testproblem $(\alpha, \omega, \Omega - \omega)$ heißt invariant bezüglich einer Gruppe G, wenn gilt:

$$\bigwedge_{g \in G} \quad \bar{g}\,\omega = \omega \qquad \text{und}$$

$$\bar{g}\,(\Omega - \omega) = \Omega - \omega .$$

Def. 10: Ein Test $\varphi : S \to [0, 1]$ heißt invariant bezüglich einer Gruppe G, wenn

$$\bigwedge_{g \in G} \bigwedge_{x \in S} \quad \varphi(gx) = \varphi(x) .$$

Invariante Tests werden natürlich überall dort bevorzugt Anwendung finden, wo ein invariantes Testproblem vorliegt. Dann wird auch die Suche nach einem gleichmäßig besten Test für das Problem $(\alpha, \omega, \Omega-\omega)$ auf invariante Tests unter allen Tests $\varphi \in M_\alpha$ beschränkt sein. Dabei kann es allerdings passieren, daß schon die Zulässigkeit im Sinne von Definition 7 nicht mehr erfüllt ist. Lehmann zeigt nämlich an einem Beispiel[1], daß nicht jeder gleichmäßig beste invariante Test

1) Lehmann, [36], S. 231

zulässig ist, was hingegen nach Satz 2 für jeden gleichmäßig besten unverfälschten Test zutrifft.

2.2 Invariante und fundamental invariante Abbildungen und ihre Anwendung auf Testprobleme

Zur näheren Charakterisierung invarianter Tests wird zunächst der Zusammenhang zwischen invarianten und fundamental invarianten Abbildungen herausgestellt.

Def. 11 : Eine Abbildung $h : S \longmapsto W$ heißt invariant bezüglich einer Gruppe G, wenn gilt :

$$\bigwedge_{x \in S} \quad \bigwedge_{g \in G} \quad h(gx) = h(x) .$$

Def. 12 : Eine invariante Abbildung $m : S \longmapsto W$ heißt fundamental invariant bezüglich einer Gruppe G, wenn aus

(3) $m(x_1) = m(x_2)$ für $x_1, x_2 \in S$ folgt:

$$\bigvee_{g \in G} \quad \text{mit} \quad x_2 = gx_1 .$$

Es bezeichne $x_1 \equiv x_2(G)$, wenn $\bigvee_{g \in G}$ mit $x_2 = gx_1$.

Da G eine Gruppe ist, erfolgt durch diese Kongruenzbeziehung eine Einteilung aller $x \in S$ in Äquivalenzklassen. Eine invariante Abbildung ist konstant auf jeder Äquivalenzklasse; eine fundamental invariante Abbildung hat wegen der Forderung (3) die zusätzliche

Eigenschaft, daß die konstanten Werte der Abbildung auf jeder Äquivalenzklasse verschieden sind.

Satz 4 :[1] Sei $m : S \to V$ fundamental invariant bezüglich einer Gruppe G und $h : S \to W$ eine beliebige Abbildung. Dann ist h bezüglich G dann und nur dann invariant, wenn eine Abbildung $f : V \to W$ existiert mit:

$$\bigwedge_{x \in S} \quad h(x) = f(m(x)) \ .$$

Beweis: Gilt $\bigwedge_{x \in S} \quad h(x) = f(m(x))$, dann ist :

$$h(gx) = f(m(gx)) = f(m(x)) = h(x) \ ,$$

d.h. h ist invariant.

Sei nun umgekehrt h invariant und $h(x) = w$ und $m(x) = v$, dann kann f definiert werden durch $f(v) = w$.
Daß f wirklich eine Abbildung ist, läßt sich wie folgt zeigen:

Ist $w_1 = h(x_1) \neq h(x_2) = w_2$, so existiert wegen der Invarianz von h kein $g \in G$ mit $x_2 = gx_1$; dann folgt aber aus der fundamentalen Invarianz von m :

$$v_1 = m(x_1) \neq m(x_2) = v_2$$

q.e.d.

Durch die Kenntnis einer bezüglich einer Gruppe G fundamental invarianten Abbildung wird also ein Überblick über alle bezüglich G

1) Lehmann, [36], S. 216

invarianten Abbildungen gewonnen. Nun soll gezeigt werden, daß sich insbesondere jeder invariante Test als Produkt aus einer meßbaren Abbildung und einer fundamental invarianten meßbaren Abbildung darstellen läßt.

Satz 5 : Für den Stichprobenraum $(S, \mathcal{A})$ sei $\varphi : S \to [0, 1]$ ein bezüglich einer Gruppe G invarianter Test und $m : S \to V$ eine bezüglich G fundamental invariante Abbildung. Weiterhin sei $\mathcal{C}$ die Menge aller meßbaren Mengen $C \subset V$, für die gilt :

$$m^{-1}(C) \in \mathcal{A} .$$

Dann kann φ durch die Abbildung $f \cdot m$ dargestellt werden, wobei f eine $\mathcal{C}$-meßbare Abbildung ist.

Beweis: Durch die Definition von $\mathcal{C}$ ist zunächst m als $(\mathcal{A}, \mathcal{C})$-meßbare Abbildung von $S \to V$ erklärt.
Nach Satz 4 gibt es eine Darstellung von φ durch $\varphi(x) = f(m(x))$ mit $f : V \to [0, 1]$.
Es bleibt zu zeigen, daß f $\mathcal{C}$-meßbar ist. Sei B eine beliebige meßbare Menge aus $[0, 1] \cap \mathcal{B}^1$.
Wegen $\varphi^{-1}(B) = m^{-1}(f^{-1}(B)) \in \mathcal{A}$ folgt notwendig $f^{-1}(B) \in \mathcal{C}$. q.e.d.

Die Bedeutung dieses Satzes liegt darin, daß mit Hilfe einer fundamental invarianten Abbildung das Testproblem vom Stichprobenraum $(S, \mathcal{A})$ auf den Meßraum $(V, \mathcal{C})$ übertragen werden kann, in dem dann oft leichter eine Lösung des Problems gefunden werden kann. Ein Beispiel dafür wird im 3. Kapitel angeführt.

Bei der Untersuchung eines Testproblems unter Berücksichtigung des Invarianz-Prinzips sind also folgende Schritte durchzuführen:

I Es ist die Gruppe G von Abbildungen anzugeben, bezüglich der das vorgelegte Testproblem invariant ist.

II Eine bezüglich G fundamental invariante Abbildung ist zu bestimmen.

III Für diese fundamental invariante Abbildung sind die induzierten W-Maße (für H_o und H_1) zu berechnen. Das Testproblem wird dann bezüglich dieser W-Maße betrachtet.

IV Unter allen Tests für dieses reduzierte Problem ist ein bester Test zu suchen.

Dieses Verfahren wird im nächsten Kapitel auf das einseitige Zwei-Stichproben-Problem angewandt werden, für das im Sinne von I eine besondere Gruppe von Abbildungen untersucht wird.

2.3 Eine spezielle Abbildungsgruppe

Es sei $(\mathbb{R}^n, \mathcal{B}^n)$ der betrachtete Stichprobenraum und G' die Gruppe aller stetigen, echt monoton wachsenden Abbildungen g' von $\mathbb{R}$ auf sich mit $g'(x_i) = x_i'$, $i = 1, \ldots, n$.

Mit Hilfe dieser Abbildungen können solche von $\mathbb{R}^n \to \mathbb{R}^n$ konstruiert werden, und zwar durch:

$$\bigwedge_{g' \in G'} \quad \bigwedge_{x \in \mathbb{R}^n} \quad gx = g(x_1, \ldots, x_n) := (g'x_1, \ldots, g'x_n) \, .$$

Die Menge aller dieser Abbildungen g bildet eine Gruppe, wenn die Verknüpfung von $g_1, g_2 \in G$ definiert wird durch

$$g_1 g_2 x := (g_1' g_2' x_1, \ldots, g_1' g_2' x_n) .$$

Das durch diese Gruppe induzierte W-Maß gemäß (2), Seite 19, ist wie folgt bestimmt:

Die Parametermenge Ω sei die Menge aller über $(\mathbb{R}^n, \mathfrak{B}^n)$ definierten Verteilungsfunktionen $F^{(n)}(x_1, \ldots, x_n)$, die wegen der vorausgesetzten Unabhängigkeit der $X_1, \ldots, X_n$ als Produkt von über $(\mathbb{R}, \mathfrak{B})$ definierten Verteilungsfunktionen $F_i(x_i)$, $i = 1, \ldots, n$, eindeutig bestimmt sind. Die zu $P_{\bar{g}F^{(n)}}$ gehörige Verteilungsfunktion $\bar{g}F^{(n)}$ läßt sich dann folgendermaßen berechnen:

Sei I^n ein Intervall des $\mathbb{R}^n$ der Form

$$\prod_{i=1}^{n} (-\infty, a_i) ,$$ dann gilt nach (2), Seite 19 :

$$P_{\bar{g}F^{(n)}}(I^n) = P_{F^{(n)}}(g^{-1} I^n) = \prod_{i=1}^{n} F_i((g')^{-1} a_i)$$

$$= \prod_{i=1}^{n} \bar{g}' F_i(a_i) , \quad \text{d.h.}$$

$$\bigwedge_{a \in \mathbb{R}^n} \bar{g}F^{(n)}(a) = \prod_{i=1}^{n} F_i((g')^{-1} a_i) .$$

Bezüglich der oben beschriebenen Gruppe G soll nun eine fundamental invariante Abbildung im Sinne von II (Seite 25) für eine Teilmenge

von $\mathbb{R}^n$ nachgewiesen werden. Dazu seien zunächst einige Bezeichnungen eingeführt:

Für jedes $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ sei mit $x^{(i)} = p_i(x)$ die i-te kleinste Komponente von x bezeichnet, $i = 1, \ldots, n$, d.h. $x^{(1)} \leq x^{(2)} \leq \ldots \leq x^{(n)}$. Bezogen auf die Zufallsvariable $X = (X_1, \ldots, X_n)$ wird $X^{(i)} = p_i(X)$ i-te geordnete Statistik genannt. Unter der Voraussetzung, daß $x_i \neq x_j$ für $i \neq j$ gilt, bezeichne $r_i(x) = r_i$ die Anzahl der Komponenten von x, die kleiner als x_i sind, d.h. $x_i = x^{(r_i)}$; r_i heißt Rang von x_i bezüglich x, und analog wird $R_i = r_i(X)$ i-te Rangstatistik genannt.

Satz 6: Es sei G die oben angegebene Gruppe und

$K := \{x = (x_1, \ldots, x_n) \in \mathbb{R}^n \mid x_i \neq x_j \text{ für } i \neq j,\ i, j = 1, \ldots, n\}$

Dann ist die Abbildung

$m : K \to \mathbb{N}^n$ mit $m(x_1, \ldots, x_n) = (r_1, \ldots, r_n)$ fundamental invariant bezüglich G.

Beweis: m ist zunächst invariant bezüglich G, denn es gilt :

$$\bigwedge_{g \in G} \bigwedge_{x \in K} m(gx) = m(g'x_1, \ldots, g'x_n) = mx \quad ,$$

da g' echt monoton wachsend ist.

Haben nun umgekehrt die Koordinaten von

$x = (x_1, \ldots, x_n)$ und $x' = (x'_1, \ldots, x'_n)$ mit $x, x' \in K$

und $x \neq x'$ dieselbe Ordnung, d.h. gilt :

$$x_{i_1} < x_{i_2} < \ldots < x_{i_n} \quad \text{und} \quad x'_{i_1} < x'_{i_2} < \ldots < x'_{i_n}$$

und damit

$$m(x_1,\ldots, x_n) = m(x'_1,\ldots, x'_n) = (r_1,\ldots, r_n) ,$$

dann existiert ein $g \in G$ mit $gx = x'$.

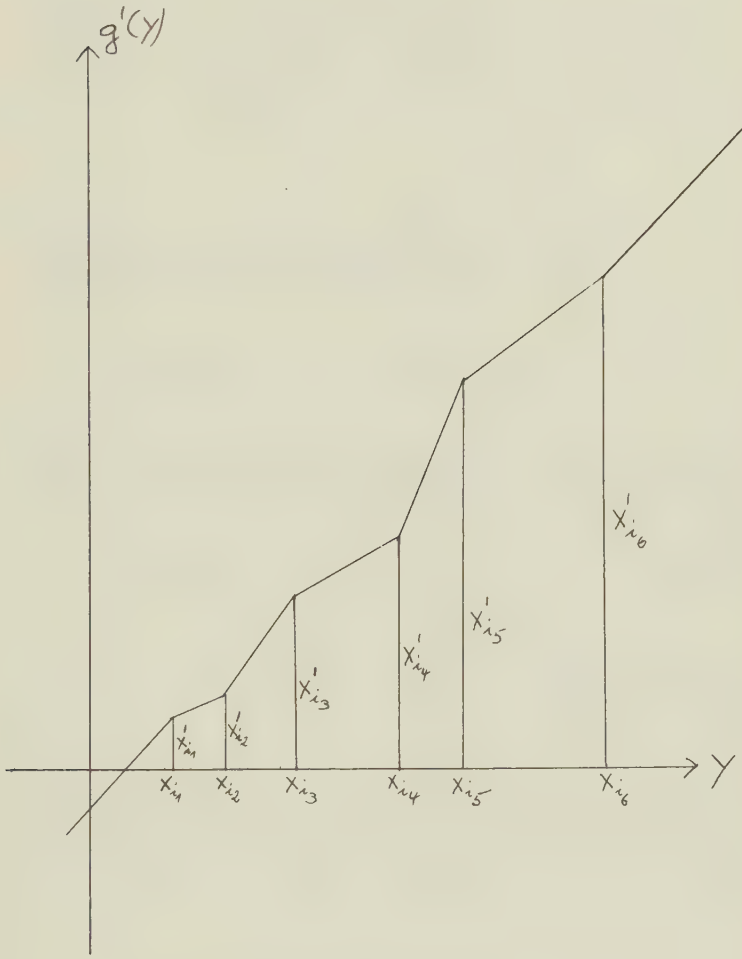
Dieses g kann nach Seite 25 definiert werden durch $g'\colon \mathbb{R} \to \mathbb{R}$ mit

$$g'(y) = \begin{cases} y + (x'_{i_1} - x_{i_1}) & \text{für } y \leq x_{i_1} \\ \dfrac{x'_{i_{j+1}} - x'_{i_j}}{x_{i_{j+1}} - x_{i_j}} (y - x_{i_j}) + x'_{i_j} & \text{für } x_{i_j} \leq y \leq x_{i_{j+1}} , \\ & j = 1,\ldots, n-1 \\ y + (x'_{i_n} - x_{i_n}) & \text{für } y \geq x_{i_n} \end{cases}$$

Es gilt nämlich: $\bigwedge_{j=1,\ldots,n} g'(x_{i_j}) = x'_{i_j}$.

Weiterhin ist g' stetig und echt monoton wachsend. Die Funktion m ist also auf K fundamental invariant bezüglich G . q.e.d.

Graphische Veranschaulichung der Abbildung g' für $j = 1,\ldots,6$:



Formal läßt sich m durch seinen Bildwert, das Tupel $(r_1, \ldots, r_n)$ bezüglich $(x_1, \ldots, x_n)$ beschreiben; die Ränge r_i sind dabei nur für alle $x \in \mathbb{R}^n - K$ nicht eindeutig definiert. Wird aber vorausgesetzt, daß die zu x_i gehörigen Stichprobenvariablen X_i alle dieselbe stetige Verteilungsfunktion F besitzen $(i = 1, \ldots, n)$, so wie es im Zusammenhang mit dem in 3.2 zu diskutierenden Zwei-Stichproben-Problem der Fall ist, dann bildet die Menge $\mathbb{R}^n - K$ eine P^X-Nullmenge.

Aus Satz 5 und Satz 6 folgt das

Korollar: Jeder bezüglich der betrachteten Gruppe G invariante Test $\varphi: K \to [0, 1]$ ist eine meßbare Abbildung des Rangtupels $(r_1, \ldots, r_n)$ von $(x_1, \ldots, x_n)$.

Diese bezüglich G invarianten Tests werden auch ranginvariant genannt.

Ein wichtiger Spezialfall ist gegeben durch die Gruppe G^* aller Abbildungen $g^*: \mathbb{R}^n \to \mathbb{R}^n$, die nach Seite 25 durch die Menge aller Abbildungen $g^{**}: \mathbb{R} \to \mathbb{R}$

mit: $g^{**}(x_i) = ax_i - b$, $a, b \in \mathbb{R}$, $a > 0$ bestimmt sei.

Wegen $G^* \subset G$ ist jedes bezüglich G invariante Testproblem auch bezüglich G^* invariant.

Ein Element $B \in \mathfrak{B}^n$, für das gilt:

$$[(x_1, \ldots, x_n) \in B] \iff [(ax_1 - b, \ldots, ax_n - b) \in B]$$

heißt dann invariant bezüglich Transformationen der "Lage" und der

"Skalierung". Alle diese Elemente erzeugen offensichtlich eine σ - Algebra $\mathcal{B}^*$ in $\mathcal{B}^n$.

Für jedes $\sigma > 0$, $\mu \in \mathbb{R}$ stimmen die beiden W-Maße P und $P_{\mu\sigma}$, die durch die Dichten

$$p = p(x_1, \ldots, x_n) \qquad \text{bzw.}$$

$$P_{\mu\sigma} = \sigma^{-n}\, p\left(\frac{x_1 - \mu}{\sigma}, \ldots, \frac{x_n - \mu}{\sigma}\right)$$

bestimmt sind, auf $\mathcal{B}^*$ überein, d.h. es gilt:

$$\bigwedge_{B \in \mathcal{B}^*} P(B) = P_{\mu\sigma}(B) \quad .$$

Ein Test φ wird dann invariant bezüglich der Gruppe der Transformationen von Lage und Skalierung genannt, wenn φ $\mathcal{B}^*$ - meßbar ist.

3. Kapitel Rangtests als Spezialfall nichtparametrischer Tests

3.1 Der Begriff "nichtparametrischer" Test

Die Definition eines nichtparametrischen Tests ist in der Literatur nicht einheitlich. So wie Fisz[1] unter einem solchen Test lediglich einen Anpassungstest versteht, der die spezielle Gestalt der Verteilungsfunktion der Grundgesamtheit prüfen soll, so beziehen sich auch die meisten anderen Lehrbücher bei der Erklärung nichtparametrischer Methoden auf spezielle Probleme. Eine allgemeine Definition - ähnlich wie die von Schmetterer[2] - führt Witting[3] an. Eine eingehende Untersuchung des Begriffs "nichtparametrisch" bringt eine Arbeit von Kendall/Sundrum[4].

Bei den parametrischen Verfahren hat die Wahrscheinlichkeitsverteilung eine funktionelle Form, und alle Hypothesen zur vollständigen Spezifizierung dieser Verteilungsfunktion beziehen sich auf eine Parametermenge $\Omega \subseteq \mathbb{R}^n$. Wird nun keine spezielle Annahme über die Gestalt der Verteilung gemacht (daher auch die Bezeichnung "verteilungsfreie" Methoden), so lassen sich die einzelnen Elemente von $\{P_\theta \mid \theta \in \Omega\}$ nicht mehr durch $\theta \in \Omega$ charakterisieren. In diesem Sinne sind nichtparametrische Tests dadurch gekennzeichnet, daß zu ihrer Anwendung alle (stetigen) Verteilungen zugelassen sind, die dann als Parameter aufgefaßt werden können. Die Parametermenge Ω ist nun der (unendlich-dimensionale) Funktionenraum aller (stetigen) Verteilungsfunktionen.

---

1) Fisz, [17], S. 364
2) Schmetterer, [58], S. 492
3) Witting, [68], S. 16
4) Kendall/Sundrum, [31], S. 124

3.2 Das Zwei-Stichproben-Problem mit einigen speziellen Alternativen

Einer der wichtigsten Anwendungsbereiche nichtparametrischer Tests bezieht sich auf das sogenannte Zwei-Stichproben-Problem.

Es seien $X_1, \ldots, X_m$ und $Y_1, \ldots, Y_n$ unabhängige Stichprobenvariable mit stetigen Verteilungsfunktionen $P(X_i \leq z) = F_1(z)$ bzw. $P(Y_j \leq z) = F_2(z)$, $i = 1, \ldots, m$ und $j = 1, \ldots, n$. [1)]

Wegen der Unabhängigkeit der X_i und Y_j gilt für die gemeinsamen Verteilungen :

$$F_1^{(m)}(x_1, \ldots, x_m) = \prod_{i=1}^{m} F_1(x_i) \qquad \text{und}$$

$$F_2^{(n)}(y_1, \ldots, y_n) = \prod_{j=1}^{n} F_2(y_j) \quad .$$

Für die kombinierte Stichprobe $x_1, \ldots, x_m, y_1, \ldots, y_n$ ist dann der zum Stichprobenraum $(\mathbb{R}^{n+m}, \mathcal{B}^{n+m})$ gehörige Wahrscheinlichkeitsraum durch das Tripel $(\mathbb{R}^{n+m}, \mathcal{B}^{n+m}, H^{(n+m)})$ mit

$$H^{(n+m)} = \frac{m}{n+m} F_1^{(m)} + \frac{n}{n+m} F_2^{(n)} \qquad \text{gegeben.}$$

Die Menge der (zulässigen) Hypothesen sei die Menge Ω aller Paare stetiger Verteilungsfunktionen (F_1, F_2), und es sei

$$\omega \subset \Omega \quad \text{mit} \quad \omega = \{ (F_1, F_2) \mid \bigwedge_{z \in \mathbb{R}} F_1(z) = F_2(z) \}.$$

---

1) Diese Voraussetzungen gelten für alle weiteren Ausführungen.

Dann kann das Zwei-Stichproben-Problem für dieses ω formuliert werden als Test von:

$$(1) \qquad H_o : (F_1, F_2) \in \omega$$

$$H_1 : (F_1, F_2) \in \Omega - \omega \ .$$

Für solche Alternativen H_1, über die also noch "völlige Unsicherheit" herrscht, existiert kein gleichmäßig bester Test.
Um konkretere Aussagen über die Testgüte machen zu können, empfiehlt es sich, als Alternativen gewisse Teilmengen von $\Omega - \omega$ auszuzeichnen. Wird z.B. vermutet, daß die Y-Werte im Durchschnitt kleiner oder größer sind als die X-Werte, so wird die Alternative festgesetzt durch:

$$(2) \qquad H_1^* : (F_1, F_2) \in \Gamma_1 \quad \text{oder} \quad H_1^{**} : (F_1, F_2) \in \Gamma_2, \ F_1 \neq F_2, \text{ mit:}$$

$$\Gamma_1 = \{ (F_1, F_2) \mid \bigwedge_{z \in \mathbb{R}} F_2(z) \geq F_1(z) \} \qquad \text{bzw.}$$

$$\Gamma_2 = \{ (F_1, F_2) \mid \bigwedge_{z \in \mathbb{R}} F_2(z) \leq F_1(z) \}$$

Die Y_j heißen stochastisch kleiner bzw. stochastisch größer als die X_i, und das Problem (2) wird als einseitiges Zwei-Stichproben-Problem bezeichnet.

Die im 4. Kapitel herzuleitenden besten und lokal besten Rangtests beziehen sich auf die folgenden aus (1) und (2) speziell gewählten Alternativen:

(3) $F_2(z) = h(F_1(z))$, wobei $h : [0, 1] \to [0, 1]$

mit $h(0) = 0$ und $h(1) = 1$ differenzierbar und monoton wachsend ist.

$H_0 : h = Id$

$H_1 : h \neq Id$.

Da die spezielle Gestalt der Verteilungsfunktionen F_1 und F_2 nicht in die Hypothese eingeht, handelt es sich hier um eine sogenannte nichtparametrische Alternative.

Beispiel:

$F_2(z) = (F_1(z))^k$, $k \in \mathbb{N}$,

$H_0 : k = 1$

$H_1 : k > 1$.

Diese Alternative kann wie folgt interpretiert werden:
Die Verteilung von Y (2. Stichprobe) ist dieselbe wie die des Maximums von k (unabhängigen) Variablen X (1. Stichprobe) :

$$F_2(z) = (F_1(z))^k = (P(X \leq z))^k = P(\max_{i=1,\ldots,k} X_i \leq z) .$$

Weiterhin gilt: $P(X < Y) = \int F_1 \, dF_2 = \frac{k}{k+1}$.

(4) $F_1(z) = K(z, \theta)$

$F_2(z) = K(z, \phi)$, $\theta, \phi \in \mathbb{R}$,

wobei K eine beliebige, aber feste Verteilungsfunktion ist.

$$H_o : \theta = \phi = : \theta_o$$

$$H_1 : \theta \neq \phi \qquad \text{(zweiseitig)}$$

$$H_1^* : \theta > \phi \quad \text{bzw.} \quad \theta < \phi \quad \text{(einseitig)} .$$

Die Entscheidung für eine solche parametrische Alternative ist auch bei nichtparametrischen Tests von Bedeutung, ermöglicht sie doch einen Vergleich dieser Tests mit den entsprechenden parametrischen Tests, wie dem t-Test und F-Test (5. Kapitel).

Die beiden wichtigsten Teilprobleme von (4) , auf die sich auch die genannten parametrischen Verfahren beziehen, beinhalten das Prüfen von Unterschieden der beiden Verteilungsfunktionen F_1 und F_2 hinsichtlich

(4a) der Lage

$$F_1(z) = K(z - \theta)$$

$$F_2(z) = K(z - \phi); \ \theta, \phi \in \mathbb{R} .$$

F_1 und F_2 unterscheiden sich also nur durch die "Lageparameter" θ und ϕ ; die Dispersionen werden als gleich vorausgesetzt.

(4b) der Variabilität

$$F_1(z) = K\left(\frac{z}{\theta}\right)$$

$$F_2(z) = K\left(\frac{z}{\phi}\right) ; \ \theta, \phi \in \mathbb{R} .$$

F_1 und F_2 unterscheiden sich also nur durch die "Dispersionsparameter" θ und ϕ ; die Mediane werden als gleich vorausgesetzt.

(5) $K(z,\phi) = (1 - \phi + \theta)\, K(z,\theta) + (\phi - \theta)\, h\left[K(z,\theta)\right]$,

wobei $h : [0,1] \to [0,1]$ mit $h(0) = 0$, $h(1) = 1$ monoton wachsend und differenzierbar ist und $K(z,\theta)$ die zu X_i und $K(z,\phi)$ die zu Y_j gehörigen Verteilungsfunktionen sind mit $\theta, \phi \in R$, $0 < \phi - \theta < 1$.

Die Verteilung der Y_j ist also ein "Gemisch" aus $K(z,\theta)$ mit dem Anteil $1 - p(p = \phi - \theta)$ und $h\left[K(z,\theta)\right]$ mit dem Anteil p. Daher wird die so formulierte Alternative, die wegen $\phi > \theta$ einseitig ist, auch Mischungsalternative genannt. Im Kapitel 4.3 wird der Spezialfall $h\left[K(z,\theta)\right] = K^2(z,\theta)$ diskutiert.

Die Menge der nichtparametrischen Tests, die mittlerweile für die verschiedensten Alternativen im Zwei-Stichproben-Problem konstruiert und auf ihre Eigenschaften untersucht worden sind, ist kaum noch überschaubar. Die meisten dieser Tests lassen sich zu den folgenden drei Gruppen zusammenfassen:

1. Tests vom Kolmogoroff-Smirnow-Typ (K.-S.-Typ)

   Dazu gehören u.a. die Tests vom Cramér-Mises und Renyi. Allen diesen Tests ist gemein, daß in die zugehörigen Teststatistiken die empirische Verteilungsfunktion eingeht.

2. Tests über Ordnungsgrößen

   Zu ihnen zählen u.a. der Median-Test, die Tests von Haga und Kamat.

3. Tests über Ränge (Rangtests)

Dazu gehören die bekannten Tests von Wilcoxon, Fischer-Yates, v.d. Waerden, Klotz, Ansari/Bradley, Mood u.a. .

Die umfangreichste Gruppe bilden zweifellos die Rangtests, die in der Praxis ihrer einfachen Anwendung wegen große Bedeutung erlangt haben. Es wird sich zeigen, daß auch von der Theorie her im Zusammenhang mit optimalen Eigenschaften die besondere Stellung der Rangtests unter den nichtparametrischen Tests gegeben ist.

3.3 Rangtests - Permutationstests

Durch das Korollar (Seite 30) wird eine bevorzugte Anwendung rang-invarianter Tests als meßbare Funktionen der Ränge für solche Testprobleme begründet, die invariant unter der Gruppe aller stetigen und echt monoton wachsenden Abbildungen sind. Aber auch für Testprobleme, bei denen die Annahme der Invarianz nicht gerechtfertigt erscheint, werden Rangtests[1] herangezogen. Diese können nicht nur als ein Spezialfall invarianter Tests betrachtet werden, sondern auch der Klasse der Permutationstests zugeordnet werden, die Funktionen der Ränge $r_1, \ldots, r_n$ und der Ordnungsgrößen $x^{(1)}, \ldots, x^{(n)}$ der Stichprobe $x_1, \ldots, x_n$ sind. Es läßt sich zeigen, daß die Klasse der Permutationstests mit der Klasse der ähnlichen Tests[2] übereinstimmt[3], ihre kritischen Bereiche somit zu einem Dynkin-System ähnlicher Ereignisse[4] gehören.

---

1) Eine exakte Definition der Rangtests im Zusammenhang mit dem Zwei-Stichproben-Problem erfolgt in 3.4.2

2) Ein Test heißt ähnlich für das Problem $(\alpha, \omega, \Omega - \omega)$, wenn gilt: $\bigwedge_{\theta \in \omega} E_\theta \varphi(X) = \alpha$, Schmetterer, [58], S. 265

3) Lehmann, [36], S. 184

4) Ein Ereignis A heißt ähnlich bezüglich einer Menge von W-Maßen $\{ P_\theta \mid \theta \in \Omega \}$, wenn gilt: $\bigwedge_{\theta \in \Omega} P_\theta(A) = \text{konst.}$.

Wesentlich ist dabei die Tatsache, daß ähnliche Ereignisse keine σ - Algebra, sondern nur ein Dynkin-System bilden. Die bedingte Wahrscheinlichkeit und der bedingte Erwartungswert (bezogen auf σ - Algebren) haben bekanntlich kein "Gegenstück" in Dynkin-Systemen, auf die sich auch der Satz von Radon-Nikodym nicht übertragen läßt. Das bedeutet, daß das Neyman-Pearson-Lemma in Dynkin-Systemen und damit für Permutationstests nicht ableitbar ist. Die Tatsache aber, daß die kritischen Bereiche von Rangtests eine σ - Algebra erzeugen, zeigt den (theoretischen) Vorteil von Rangtests gegenüber Permutationstests. Der praktische Vorteil ist wegen des weit geringeren Rechenaufwandes evident.

3.4 Anwendung des Invarianz-Prinzips auf das einseitige Zwei-Stichproben-Problem

3.4.1 Herleitung einer suffizienten Rangstatistik

Nach dem auf Seite 25 beschriebenen Verfahren I - IV soll nun das einseitige Zwei-Stichproben-Problem, hier kurz dargestellt durch:

$$H_o : F_1 = F_2 \qquad \text{für alle } z \in \mathbb{R}$$

$$H_1 : F_1 \geq F_2 \quad [F_1 \leq F_2] \quad \text{für alle } z \in \mathbb{R} \quad \text{und} \quad F_1 \neq F_2 ,$$

schrittweise reduziert werden:

I Das Problem bleibt invariant unter der Gruppe G, deren Elemente g definiert sind durch

$$g(x_1, \ldots, x_n) := (g'x_1, \ldots, g'x_n), \text{ wobei } g' : \mathbb{R} \to \mathbb{R}$$

stetig und echt monoton wachsend ist (Seite 25). Sind nämlich

$X_1, \ldots, X_m$ und $Y_1, \ldots, Y_n$ Stichprobenvariable, in denen die X_i und Y_j unabhängig sind und stetige Verteilungsfunktionen F_1 bzw. F_2 haben, dann sind einerseits auch

$$X_i' := g'(X_i) \qquad \text{bzw.} \qquad Y_j' := g'(Y_j)$$

$i = 1, \ldots, m$, $j = 1, \ldots, n$ unabhängige Zufallsvariable mit stetigen Verteilungsfunktionen. Andererseits sind auch die Hypothesen invariant, denn es gilt für die durch g induzierte Abbildung $\bar{g}$ (Seite 26):

$$\bar{g}F_1(z) = F_1((g')^{-1}(z)) \geq F_2((g')^{-1}(z)) = \bar{g}F_2(z) \ .$$

II Die zu bestimmende fundamental invariante Abbildung l bezüglich der Gruppe G ist nach Satz 6 gegeben durch:

$$(6) \quad l(x_1,\ldots, x_m, y_1,\ldots, y_n) = (s_1^*,\ldots, s_m^*, t_1^*,\ldots, t_n^*),$$

wobei die s_i^* und die t_j^* die Ränge von x_i bzw. y_j in der kombinierten Stichprobe darstellen.

Neben dieser Reduktion durch Invarianz, die die Menge der Entscheidungsfunktionen reduziert, gelingt eine weitere Restriktion des Testproblems durch Suffizienz, d.h. eine Verkleinerung der σ - Algebra des Stichprobenraumes.

Es sei $\mathbf{L} : \mathbb{N}^{n+m} \to \mathbb{N}^{n+m}$ mit

$L(s_1^*,\ldots, s_m^*, t_1^*,\ldots, t_n^*) = (s_1,\ldots, s_m, t_1,\ldots, t_n)$,

wobei die s_i und t_j die geordneten Ränge der x_i bzw. y_j bedeuten, d.h. $s_1 < s_2 < \ldots < s_m$, $t_1 < t_2 < \ldots < t_n$.

Dann ist L suffizient für die Menge aller W-Maße $P_{F_1F_2}$, die durch $l(x_1, \ldots, x_m, y_1, \ldots, y_n) = (s_1^*, \ldots, s_m^*, t_1^*, \ldots, t_n^*)$ induziert werden. Es gilt nämlich:

$$P_{F_1F_2}(S_1^* = s_1^*, \ldots, S_m^* = s_m^*, T_1^* = t_1^*, \ldots, T_n^* = t_n^* \mid L) = \frac{1}{m!n!},$$

da die Verteilungen von $(S_1^*, \ldots, S_m^*, T_1^*, \ldots, T_n^*)$ invariant sind gegenüber den $m!$ Permutationen innerhalb der ersten m Komponenten und dem $n!$ Permutationen der letzten n Komponenten. Weil weiterhin die Ränge $s_1, \ldots, s_m$ eindeutig die Ränge $t_1, \ldots, t_n$ bestimmen und umgekehrt (jeweils als Komplementärmenge von $\{1, \ldots, n+m\}$), kann man sich auf eine der beiden suffizienten Statistiken L_1 oder L_2 mit

$$(7) \qquad L_1(s_1^*, \ldots, s_m^*, t_1^*, \ldots, t_n^*) = (s_1, \ldots, s_m) \qquad \text{bzw.}$$

$$(7^*) \qquad L_2(s_1^*, \ldots, s_m^*, t_1^*, \ldots, t_n^*) = (t_1, \ldots, t_n)$$

beschränken; so im folgenden auf L_2.

Damit wird der ursprüngliche Stichprobenraum $(\mathbb{R}^{n+m}, \mathfrak{B}^{n+m})$ reduziert auf $(T, \mathcal{P})$, wobei

$$T = \{(t_1, \ldots, t_n) \mid t_i \in \mathbb{N},\ t_i < t_{i+1},\ i = 1, \ldots, n-1\}$$

ist und $\mathcal{P}$ die Potenzmenge von T.

Jede Abbildung $\varphi : T \to [0,1]$ kann dann als Rangtest für das einseitige Zwei-Stichproben-Problem betrachtet werden, da φ $\mathcal{P}$- meßbar ist.

Die meisten bekannten Rangtests für das obige Problem werden durch Teststatistiken der Form $\sum_{j=1}^{n} h(t_j)$ mit $h : \mathbb{N} \to \mathbb{R}$ beschrieben.

3.4.2 Verteilung der Ränge unter $F_1 = F_2$ und $F_1 \neq F_2$

III Nach den Ausführungen in II wird nun das Testproblem untersucht für die W-Maße

$P_{F_1F_1}(T_1 = t_1, \ldots, T_n = t_n)$ (Nullhypothese : $F_1 = F_2$)

$P_{F_1F_2}(T_1 = t_1, \ldots, T_n = t_n)$ (Alternativhypothese: $F_1 \neq F_2$)

mit $1 \leq t_1 < t_2 < \ldots < t_n \leq N$ und $N = n + m$.

$P_{F_1F_1}$ ist leicht zu bestimmen. Da T genau $\binom{N}{n}$ Elemente enthält und die Wahrscheinlichkeit für das Auftreten eines jeden Elementes $(t_1, \ldots, t_n) \in T$ unter H_o gleich ist, gilt also

(8) $$P_{F_1F_1}(T_1 = t_1, \ldots, T_n = t_n) = \frac{1}{\binom{N}{n}}$$

für alle t_i mit $1 \leq t_1 < t_2 < \ldots < t_n \leq N$,

d.h. $P_{F_1F_1}$ ist unabhängig von F_1 .

Jeder diesbezügliche Rangtest vom Niveau $\alpha = \frac{q}{\binom{N}{n}}$ hat demnach einen kritischen Bereich C, der aus genau q Elementen $(t_1,\ldots, t_n) \in T$ besteht. Gilt für beliebiges α mit $0 < \alpha < 1$: $\alpha \cdot \binom{N}{n} \notin \mathbb{N}$, so wird $q^* = \left[\alpha \cdot \binom{N}{n}\right]$ gesetzt.

Die Bestimmung von $P_{F_1F_2}$ gelingt über folgenden Satz von Hoeffding:

Satz 7:[1] (i) Es seien $V_1,\ldots, V_N$ unabhängige Zufallsvariable, V_i besitze eine Verteilung mit der Dichte f_i und R_i bezeichne den Rang von V_i, $i = 1,\ldots, N$. Weiterhin seien $W_1,\ldots, W_N$ unabhängige Zufallsvariable, wobei alle W_i eine Verteilung mit der Dichte f haben.

(ii) Die Verteilung von V_i sei total stetig bezüglich der von W_i. Dann gilt:

$$(9) \qquad P(R_1 = r_1,\ldots, R_N = r_N) = \frac{1}{N!}\, E\left[\prod_{i=1}^{N} \frac{f_i(W^{(r_i)})}{f(W^{(r_i)})}\right].$$

Beweis: Es gilt $P(R_1 = r_1,\ldots, R_N = r_N) =$

$$\int \ldots \int f_1(v_1)\ldots f_N(v_N)\, dv_1 \ldots dv_N ,$$

integriert über einen Bereich, in dem v_i der r_i- kleinste der $v_1, \ldots, v_N$ ist, $i = 1, \ldots, N$.

---

1) Lehmann, [36], S. 254

Mit Hilfe von $z_{r_i} := v_i$ ergibt sich dann:

$$P(R_1 = r_1, \ldots, R_N = r_N) =$$

$$= \int\limits_{-\infty<z_1<\ldots<z_N<\infty} \ldots \int f_1(z_{r_1}) \ldots f_N(z_{r_N}) \cdot dz_1, \ldots, dz_N \ .$$

Daraus folgt wegen (ii) :

$$P(R_1 = r_1, \ldots, R_N = r_N) =$$

$$= \frac{1}{N!} \int\limits_{-\infty<z_1<\ldots<z_N<\infty} \ldots \int \frac{\prod_{i=1}^{N} f_i(z_{r_i})}{\prod_{i=1}^{N} f(z_i)} \, N! \prod_{i=1}^{N} f(z_i)\, dz_i \ .$$

Da die Dichte h von $\left(W^{(1)}, \ldots, W^{(N)}\right)$ die Gestalt hat[1)]

$$h(z_1, \ldots, z_N) = \begin{cases} N!\, f(z_1) \ldots f(z_N) & \text{für } -\infty<z_1<\ldots<z_N<\infty \\ 0 \quad \text{sonst} & , \end{cases}$$

folgt unmittelbar die Gleichung (9) . q.e.d.

Es sei nun speziell $N = n + m$, $f_1 = f_2 \ldots = f_m = f$

$f_{m+1} = \ldots = f_N = \tilde{f}$ und $V_i = X_i$, $i = 1, \ldots, m$

$V_{m+j} = Y_j$, $j=1,\ldots,n$; d.h. $(V_1,\ldots,V_N) = (X_1,\ldots,X_m, Y_1,\ldots,Y_n)$.

Mit F_1 sei die Verteilung von X_i, mit F_2 die von Y_j und mit $t_1, \ldots, t_n$ seien die geordneten Ränge von $y_1, \ldots, y_n$ in der kombinierten Stichprobe $x_1, \ldots, x_m, y_1, \ldots, y_n$ bezeichnet. Weiterhin seien die Voraussetzungen von Satz 7 er-

---

1) Schmetterer, [58], S. 501

füllt. Dann folgt unmittelbar das

Korollar:

$$P_{F_1F_2}(T_1 = t_1,\ldots,T_n = t_n) = n!m!\ \frac{1}{N!}\ E_{F_1F_1}\left[\prod_{j=1}^{n}\frac{\varphi\left(W^{(t_j)}\right)}{f\left(W^{(t_j)}\right)}\right] =$$

(10)

$$= \frac{1}{\binom{N}{n}}\ E_{F_1F_1}\left[\prod_{j=1}^{n}\frac{\varphi\left(W^{(t_j)}\right)}{f\left(W^{(t_j)}\right)}\right] \quad ,^{1)}$$

wobei $W^{(i)}$ die Verteilung F_1 hat $(i=1,\ldots,N)$.

Damit ist also das gesuchte W-Maß $P_{F_1F_2}$ bestimmt.

IV Ein bester Rangtest φ unter allen Tests für eine einfache Null- und Alternativhypothese (F_1 und F_2 sind also speziell gewählt und bestimmen vollständig $P_{F_1F_2}$) ist nach dem Neyman-Pearson-Lemma wegen $P_{F_1F_1}$ = konst. gegeben durch:

$$\varphi(t_1,\ldots,t_n) = \begin{cases} 1 \text{ für } P_{F_1F_2}(T_1 = t_1,\ldots,T_n = t_n) > a \\ b \text{ für } \quad " \quad = a \\ o \text{ für } \quad " \quad < a \end{cases} .$$

Die Bestimmung von φ ist allerdings nur selten möglich, da sich die Berechnung von

$$E_{F_1F_1}\left[\prod_{j=1}^{n}\frac{\varphi\left(W^{(t_j)}\right)}{f\left(W^{(t_j)}\right)}\right]$$

meist als zu kompliziert erweist.

---

1) Der Index F_1F_1 von E soll andeuten, daß der Erwartungswert unter $F_1 = F_2$ betrachtet wird.

3.5 Ein bester Rangtest[1] für $F_2 = F_1^k$

Im folgenden soll nun mit Hilfe des Korollars zu Satz 7 über die Alternative (3) aus 3.2 : $F_2 = h(F_1)$ für den Spezialfall $F_2 = F_1^k$ ein bester Rangtest hergeleitet werden.

Es seien f und $\tilde{f}$ die Dichten von F_1 bzw. $F_2 = h(F_1)$.

Dann gilt:

$h'(F_1(z)) = \frac{\tilde{f}(z)}{f(z)}$ und damit nach (10) (da die Voraussetzungen des Satzes 7 erfüllt sind) :

$$(11) \quad P_{F_1F_2}(T_1 = t_1,\ldots,T_n = t_n) = \frac{1}{\binom{N}{n}} E_{F_1F_1}\left[\prod_{j=1}^{n} h'\left(F_1\left(W^{(t_j)}\right)\right)\right] .$$

Da die $W^{(i)}$, $i = 1, \ldots, N$, die stetige Verteilungsfunktion F_1 haben, ist $F_1(W^{(i)})$ gleichverteilt über $(0,1)$, und es folgt:

$$P_{F_1F_2}(T_1 = t_1,\ldots,T_n = t_n) = \frac{1}{\binom{N}{n}} E\left[\prod_{j=1}^{n} h'\left(U^{(t_j)}\right)\right] ,$$

wobei $U^{(i)}$ über $(0,1)$ gleichverteilt ist $(i = 1, \ldots, N)$.

Ist nun insbesondere $h(F_1) = F_1^k$, so folgt:

$$(12) \quad P_{F_1F_1^k}(T_1 = t_1,\ldots,T_n = t_n) = \frac{k^n}{\binom{N}{n}} E\left[\left(\prod_{j=1}^{n} U^{(t_j)}\right)^{k-1}\right] .$$

Um diesen Ausdruck zu berechnen, muß die gemeinsame Verteilung von $\left(U^{(t_1)},\ldots, U^{(t_n)}\right)$ bestimmt werden.

---

1) nach Lehmann, [35], S. 23

Setzt man $t_o = 0$, $t_{n+1} = N+1$, $u_o = o$ und $u_{n+1} = 1$, so ist die gemeinsame Dichte $h(u_1,\ldots,u_n)$ von $\left(U^{(t_1)},\ldots,U^{(t_n)}\right)$ gegeben durch [1]:

$$h(u_1,\ldots,u_n) = \begin{cases} \dfrac{N!}{\prod_{j=0}^{n}(t_{j+1}-t_j-1)!} \prod_{j=0}^{n}(u_{j+1}-u_j)^{t_{j+1}-t_j-1} & \text{für } 0=u_o\leq\ldots\leq u_{n+1}=1 \\ 0 \qquad \text{sonst} . \end{cases}$$

Es sei nun

$$v_j := \frac{u_j}{u_{j+1}} \quad , j = 1, \ldots, n, \text{ d.h.}$$

$$u_j = v_j v_{j+1} \cdots v_n .$$

Da die Jacobische Determinante J bei dieser Variablentransformation die Gestalt hat : $J = v_2 v_3^2 \ldots v_n^{n-1}$, ist die zu $(V_1, \ldots, V_n)$ gehörige Dichte bestimmt durch [2]:

$$\frac{N!}{\prod_{j=0}^{n}(t_{j+1}-t_j-1)!} \prod_{j=1}^{n}(1-v_j)^{t_{j+1}-t_j-1} \prod_{j=0}^{n-1}(v_{j+1}\cdots v_n)^{t_{j+1}-t_j-1} .$$

$$\cdot (v_2 v_3^2 \ldots v_n^{n-1}) = \frac{N!}{\prod_{j=0}^{n}(t_{j+1}-t_j-1)} \prod_{j=1}^{n}(1-v_j)^{t_{j+1}-t_j-1} \prod_{j=1}^{n} v_j^{t_j-1} ,$$

d.h. die gemeinsame Dichte h^* von $(V_1, \ldots, V_n)$ ist :

---

1) Schmetterer, [58], S. 500

2) wobei entsprechend $v_o = o$ und $v_{n+1} = 1$ gesetzt sind.

$$h^*(v_1,\ldots,v_n) = \begin{cases} \dfrac{N!}{\prod_{j=0}^{n}(t_{j+1}-t_j-1)} \prod_{j=1}^{n} v_j^{t_j-1}(1-v_j)^{t_{j+1}-t_j-1} & \text{für } 0\le v_j\le 1,\ j=1,\ldots,n \\ 0 \qquad \text{sonst} \end{cases}$$

Offensichtlich sind $V_1, \ldots, V_n$ unabhängig und V_j hat eine Beta-Verteilung mit den Parametern t_j und $t_{j+1} - t_j$.

Wegen $U^{(t_1)} \ldots U^{(t_n)} = V_1 V_2^2 \ldots V_n^n$ folgt aus (12):

$$(13)\quad P_{F_1F_1^k}(T_1 = t_1,\ldots, T_n = t_n) = \frac{k^n}{\binom{N}{n}} \prod_{j=1}^{n} E\left[V_j^{jk-j}\right]$$

$$= \frac{k^n}{\binom{N}{n}} \prod_{j=1}^{n} \frac{\Gamma(t_j+jk-j)}{\Gamma(t_j)} \cdot \frac{\Gamma(t_{j+1})}{\Gamma(t_{j+1}+jk-j)}\ ,$$ 1)

wobei Γ die Gamma-Funktion bedeutet.

Damit ergibt sich:

Ein bester Rangtest φ für $H_o : F_1 = F_2$ gegen $H_1 : F_2 = F_1^k$, $k\in\mathbb{N}$, $k > 1$ ist bestimmt durch $\varphi(t_1, \ldots, t_n) =$

$$\begin{cases} 1 & \text{für } \prod_{j=1}^{n} \dfrac{\Gamma(t_j + jk - 1)\cdot\Gamma(t_{j+1})}{\Gamma(t_j)\cdot\Gamma(t_{j+1} - jk - 1)} > a \\ b & \text{für } \quad " \quad = a \\ 0 & \text{für } \quad " \quad < a \end{cases}.$$

Ist insbesondere $k = 2$, d.h. $F_2 = F_1^2$, dann folgt aus (13):

1) Fisz, [17], S. 138

$$P_{F_1F_1^2}(T_1=t_1,\ldots,T_n=t_n) = \frac{2^n}{\binom{N}{n}} \prod_{j=1}^{n} \frac{t_j\,(t_j + j - 1)\ldots(t_j + 1)}{t_{j+1}\,(t_{j+1} + j - 1)\ldots(t_{j+1} + 1)}$$

$$= \frac{2^n}{\binom{N}{n}} \cdot \frac{t_1\,(t_2 + 1)\ldots(t_n + n - 1)}{(n + m + 1)(n + m + 2)\ldots(n + 2m)} \quad .$$

Damit ist

$$\varphi^*(t_1,\ldots,t_n) = \begin{cases} 1 & \text{für } t_1(t_2 + 1)\ldots(t_n + n - 1) > a \\ b & \text{für } \quad " \quad = a \\ 0 & \text{für } \quad " \quad < a \end{cases}$$

bester Rangtest für $H_o : F_1 = F_2$ gegen $H_1 : F_2 = F_1^2$.

4. Kapitel Lokal beste Rangtests

4.1 Definition und Nachweis der Existenz lokal bester Rangtests für parametrische Alternativen

Wie bereits erwähnt, können lokal beste Tests als gleichmäßig beste Tests für Alternativen in einer hinreichend kleinen Umgebung der Nullhypothese gekennzeichnet werden. Die exakte Definition eines lokal besten Tests setzt die Definition einer Metrik über der Menge M_1 aller W-Maße P zum Stichprobenraum $(S, \mathcal{A})$ voraus. Eine solche Metrik d_1 ist z.B. gegeben durch:

(1) $d_1 : M_1 \times M_1 \to [0,1]$ mit

$$d_1(P_1,P_2) := \sup_{A \in \mathcal{A}} | P_1(A) - P_2(A) |; \; P_1,P_2 \in M_1 .$$

Ist nun wieder speziell $(\mathbb{R}^n, \mathcal{B}^n)$ der Stichprobenraum von $X = (X_1, \ldots, X_n)$ und haben alle unabhängigen Stichprobenvariablen X_i, $i = 1, \ldots, n$, dieselbe Verteilungsfunktion F_θ, $\theta \in \mathbb{R}^p$, so ist die gemeinsame Verteilung $F_\theta^{(n)}$ durch

$$F_\theta^{(n)}(x_1, \ldots, x_n) = \prod_{i=1}^{n} F_\theta(x_i) \quad \text{bestimmt.}$$

Da zu jedem $\theta \in \mathbb{R}^p$ genau ein F_θ gehört, kann über der Menge M der W-Maße für $(\mathbb{R}^n, \mathcal{B}^n)$ folgende Metrik definiert werden:

(2) $d : M \times M \Longrightarrow \mathbb{R}^+$ mit

$$d\left(F_{\theta_1}^{(n)}, F_{\theta_2}^{(n)}\right) := \left\| \theta_1 - \theta_2 \right\| , \quad \theta_1, \theta_2 \in \mathbb{R}^p .$$

Diese Metrik wird im folgenden bei der Untersuchung lokaler Gütekriterien von Rangtests verwendet, und zwar für das Zwei-Stichproben-Problem bezüglich der parametrischen Alternativen (4), (4a), (4b), (5) in 3.2.

Es seien also $X_1, \ldots, X_m$ und $Y_1, \ldots, Y_n$ unabhängige Stichprobenvariable; alle X_i, $i = 1, \ldots, m$ bzw. alle Y_j, $j = 1, \ldots, n$ haben dieselbe Verteilungsfunktion $F_1(z)$ bzw. $F_2(z)$, und es gelte:

$$F_1(z) = K(z, \theta)$$
$$F_2(z) = K(z, \phi) ,$$

wobei θ und $\phi \in \mathbb{R}$ beliebig aber fest gewählt sind und $K(z, \theta)$ und $K(z, \phi)$ für jedes θ bzw. ϕ vollständig bestimmt sind.

Es sei R_α die Menge aller Rangtests zum Niveau α für

$$H_o : \theta = \phi = : \theta_o \quad \text{gegen}$$
$$H_1 : \theta < \phi \quad [\theta > \phi] .$$

Jeder Rangtest $\varphi \in R_\alpha$ kann als Funktion der geordneten Ränge t_j der y_j, $j = 1, \ldots, n$, in der kombinierten Stichprobe $x_1, \ldots, x_m$, $y_1, \ldots, y_n$ dargestellt werden $(m + n = N)$.

Def. 13 : Sei $\beta_\varphi(\theta, \phi, N)$ die Güte von $\varphi \in R_\alpha$; $\varphi^* \in R_\alpha$ heißt lokal bester Rangtest für H_o gegen H_1, wenn

$$\bigwedge_{\varphi \in R_\alpha} \bigwedge_{N \in \mathbb{N}} \bigvee_{\varepsilon(N)>0} \bigwedge_{\|(\theta,\phi)-(\theta_o,\theta_o)\| < \varepsilon(N)} \beta_{\varphi^*}(\theta,\phi,N) \geq \beta_\varphi(\theta,\phi,N) .$$

Offensichtlich ist jeder lokal beste Rangtest zulässig im Sinne von Definition 7 in 1.2.

Satz 8 :[1] Es seien $X_1, \ldots, X_m$ und $Y_1, \ldots, Y_n$ unabhängige Stichprobenvariable und alle X_i, $i = 1, \ldots, m$ bzw. Y_j, $j = 1, \ldots, n$ haben die Verteilungsfunktion

$$F_1(z) = K(z, \theta) \quad \text{bzw.}$$

$$F_2(z) = K(z, \phi) \;, \quad \theta, \phi \in \mathbb{R} \;.$$

Folgende Bedingungen seien erfüllt:

(i) $K(z, \phi)$ besitzt eine Dichte $k(z, \phi)$ und $\bigvee_{\delta>0}$, so daß sowohl $k(z, \phi)$ als auch $k'(z, \phi) := \frac{\partial k(z,\phi)}{\partial \phi}$ stetig sind für alle $z \in \mathbb{R}$ und alle ϕ mit $| \phi - \theta_o | < \delta$.

Weiterhin existieren zwei Lebesgue-integrierbare Funktionen $L_1(z)$ und $L_2(z)$, so daß für alle $z \in \mathbb{R}$ und alle ϕ mit $| \phi - \theta_o | < \delta$ gilt :

$| k(z, \phi) | \leq L_1(z)$ und $| k'(z, \phi) | \leq L_2(z)$.

(ii) Die Verteilung von X_i ist total stetig bezüglich der von Y_j und umgekehrt, $i = 1, \ldots, m$, $j = 1, \ldots, n$.

Dann gilt:

Der lokal beste Rangtest $\varphi^*(t_1, \ldots, t_n)$ für $\theta = \phi = \theta_o$ gegen $\phi > \theta$ $[\phi < \theta]$ ist bestimmt durch :

$$\varphi^*(t_1, \ldots, t_n) = \begin{cases} 1 & \text{für } \sum_{j=1}^{n} c_{t_j} > a \quad [< a] \\ b & \text{für } \quad " \quad = a \\ 0 & \text{für } \quad " \quad < a \quad [> a] \end{cases}$$

---

1) Capon, [10], S. 88

mit $c_{t_j} = E_{\theta_o\theta_o}\left[\left(\ln k\left(W^{(t_j)},\phi\right)\right)'(\theta_o)\right]$,

wobei $W^{(i)}$ die Verteilung $K(z, \theta_o)$ hat, $i = 1, \ldots, N$, mit $N = n + m$.

Beweis: Der Nachweis der Existenz eines lokal besten Rangtests für das angegebene Problem wird mit Hilfe des Neyman-Pearson-Lemmas geführt. Dazu müssen die Verteilungen der geordneten Ränge t_j von y_j ; $j = 1, \ldots, n$ in der kombinierten Stichprobe $x_1, \ldots, x_m$, $y_1, \ldots, y_n$ sowohl unter $H_o : \theta = \phi = \theta_o$ als auch unter $H_1 : \phi > \theta \left[\phi < \theta\right]$ angegeben werden.

Es gilt:[1)]

$$(3) \qquad P_{\theta_o\theta_o}(T_1 = t_1, \ldots, T_n = t_n) = \frac{1}{\binom{N}{n}} \quad \text{und}$$

wegen des Korollars zu Satz 7 (Seite 45) :

$$(4) \qquad P_{\theta\phi}(T_1=t_1,\ldots,T_n=t_n) = \frac{1}{\binom{N}{n}} E_{\theta\theta}\left[\prod_{j=1}^{n} \frac{k\left(W^{(t_j)},\phi\right)}{k\left(W^{(t_j)},\theta\right)}\right]$$

$$= \frac{1}{\binom{N}{n}} E_{\phi\phi}\left[\prod_{i=1}^{n} \frac{k\left(W^{(s_i)},\theta\right)}{k\left(W^{(s_i)},\phi\right)}\right] ,$$

wobei die s_i die geordneten Ränge der x_i in der kombinierten Stichprobe $x_1, \ldots, x_m$, $y_1, \ldots, y_n$ bedeuten.
Das letzte Gleichheitszeichen gilt wegen

$$P_{\theta\phi}(T_1=t_1,\ldots,T_n=t_n) = P_{\theta\phi}(S_1=s_1,\ldots,S_m=s_m) .$$

---

1) Statt $P_{F_1F_1}$ und $P_{F_1F_2}$ werden im folgenden die entsprechenden Bezeichnungen $P_{\theta_o\theta_o}$ bzw. $P_{\theta\phi}$ gewählt.

Aus (i) folgt, daß $P_{\theta\phi}$ für $\theta_o - \delta < \theta\,,\ \phi < \theta_o + \delta$ stetige partielle Ableitungen hat.[1)]

Somit gilt:

$$(5)\qquad P_{\theta\phi} = P_{\theta_o\theta_o} + (\theta - \theta_o)\ \frac{\partial P_{\theta\phi}(\theta_o,\theta_o)}{\partial\theta} + (\phi - \theta_o)\ \frac{\partial P_{\theta\phi}(\theta_o,\theta_o)}{\partial\phi}$$

$$+ \|(\theta,\phi) - (\theta_o,\theta_o)\|\,\mathcal{O}(1)\,.$$

Um mit Hilfe von (4) die partiellen Ableitungen von $P_{\theta\phi}$ an der Stelle $\theta = \phi = \theta_o$ zu bestimmen, kann wegen (i) partielle Differentation und Erwartungswert vertauscht werden.[2)] D.h.:

$$(6)\qquad \frac{\partial P_{\theta\phi}(\theta_o,\theta_o)}{\partial\phi} = \frac{1}{\binom{N}{n}}\ E_{\theta_o\theta_o}\left[\sum_{j=1}^{n}\ \frac{k'\left(W^{(t_j)},\theta_o\right)}{k\left(W^{(t_j)},\theta_o\right)}\right]$$

$$= \frac{1}{\binom{N}{n}}\ \sum_{j=1}^{n}\ E_{\theta_o\theta_o}\ \left[\ln k\left(W^{(t_j)},\phi\right)'(\theta_o)\right]$$

$$=: \frac{1}{\binom{N}{n}}\ \sum_{j=1}^{n}\ c_{t_j}\ ;$$

und analog :

$$(7)\qquad \frac{\partial P_{\theta\phi}(\theta_o,\theta_o)}{\partial\theta} = \frac{1}{\binom{N}{n}}\ \sum_{i=1}^{m}\ E_{\theta_o\theta_o}\ \left[\left(\ln k\left(W^{(s_i)},\theta\right)\right)'(\theta_o)\right]$$

$$=: \frac{1}{\binom{N}{n}}\ \sum_{i=1}^{m}\ c_{s_i}\ .$$

---

1) Schmetterer, [58], S. 13/14

2) Schmetterer, [58], S. 13/14

(3), (6) und (7) in (5) eingesetzt, ergibt:

$$P_{\theta\phi} = \frac{1}{\binom{N}{n}} \left[1+(\theta-\theta_o) \sum_{i=1}^{m} c_{s_i} + (\phi-\theta_o) \sum_{j=1}^{n} c_{t_j} + \| (\theta,\phi) - (\theta_o,\theta_o) \| \mathcal{O}(1) \right].$$

Wegen $\sum_{i=1}^{m} c_{s_i} + \sum_{j=1}^{n} c_{t_j} = \sum_{h=1}^{N} c_h$ folgt :

$$P_{\theta\phi} = \frac{1}{\binom{N}{n}} \left[1+(\theta-\theta_o)\left(\sum_{h=1}^{N} - \sum_{j=1}^{n} c_{t_j} \right) + (\phi-\theta_o) \sum_{j=1}^{n} c_{t_j} + \right.$$

$$\left. + \| (\theta,\phi) - (\theta_o,\theta_o) \| \mathcal{O}(1) \right]$$

$$= \frac{1}{\binom{N}{n}} \left[1+(\phi-\theta) \sum_{j=1}^{n} c_{t_j} + (\theta-\theta_o) \sum_{h=1}^{N} c_h + \| (\theta,\phi) - (\theta_o,\theta_o) \| \mathcal{O}(1) \right].$$

Der Ausdruck $\sum_{h=1}^{N} c_h$ hängt zwar von N, nicht aber von den Rängen t_j der y_j ab, so daß er bei der Herleitung eines besten Rangtests für das spezielle Testproblem als Konstante aufgefaßt werden kann.

Nach dem Neyman-Pearson-Lemma ist der beste Rangtest φ^* unter allen $\varphi \in R_\alpha$ für

$H_o : \theta = \phi = \theta_o$ gegen

$H_1 : \phi \neq \theta$ [θ , ϕ fest]

bestimmt durch den kritischen Bereich:

$$1+(\phi-\theta) \sum_{j=1}^{n} c_{t_j} + (\theta-\theta_o) \sum_{h=1}^{N} c_h + \| (\theta,\phi) - (\theta_o,\theta_o) \| \mathcal{O}(1) > a_1 ,$$

und damit durch

$$\sum_{j=1}^{n} c_{t_j} + \mathcal{O}(1) > a \quad \text{für} \quad \phi > \theta$$

$$\sum_{j=1}^{n} c_{t_j} + \mathcal{O}(1) < a \quad \text{für} \quad \phi < \theta \ .$$

Da $P_{\theta\phi}$ als Funktion von $t_1, \ldots, t_n$ nur endlich viele Werte annehmen kann, nämlich $\binom{N}{n}$, ist der lokal beste Rangtest φ^* bestimmt durch

$$\sum_{j=1}^{n} c_{t_j} > a \quad \text{für} \quad \phi > \theta \quad \text{und}$$

$$\sum_{j=1}^{n} c_{t_j} < a \quad \text{für} \quad \phi < \theta \ . \qquad \text{q.e.d.}$$

Da die Variablen c_{t_j} von der speziellen Verteilungsfunktion K abhängen, liefert φ^* keinen gleichmäßig besten Rangtest für das allgemeine einseitige Zwei-Stichproben-Problem, wohl aber nach Satz 8 einen solchen in einer hinreichend kleinen $\varepsilon(N)$ - Umgebung für ein spezielles K aus der betrachteten Alternative, vorausgesetzt, daß der betreffende Erwartungswert existiert. Beispiele dazu folgen im nächsten Abschnitt.

Die oft wesentlich leichtere Bestimmung von lokal besten Rangtests gegenüber besten Tests für einfache Null- und Alternativhypothesen liegt darin begründet, daß sich die Statistik $\sum_{j=1}^{n} c_{t_j}$ des lokal

besten Rangtests als Erwartungswert einer Summe und nicht wie die eines besten Rangtests als Erwartungswert eines Produkts stochastisch abhängiger Zufallsvariablen berechnet.

4.2 Lokal beste Rangtests auf Lage und Variabilität, Beispiele

Mit Hilfe von Satz 8 werden nun lokal beste Rangtests für die beiden Spezialfälle des einseitigen Zwei-Stichproben-Problems (4a) und (4b) in 3.2 hergeleitet.

Korollar 1 (Lagealternativen)

Die Voraussetzungen des Satzes 8 seien für die Verteilungsfunktionen

$$F_1(z) = K(z - \theta) \qquad \text{und}$$

$$F_2(z) = K(z - \phi) \quad \text{mit den Dichten } k(z - \theta) \text{ bzw.}$$

$k(z - \phi)$ erfüllt. Dann gilt:

Der lokal beste Rangtest für

$$H_o : \theta = \phi =: \theta_o \qquad \text{gegen}$$

$$H_1 : \phi > \theta \ (\phi < \theta)$$

wird beschrieben durch die Statistik:

$$L = -\sum_{j=1}^{n} E_{\theta_o \theta_o} \left[\frac{\partial}{\partial W^{(t_j)}} \ln k\left(W^{(t_j)} \right) \right] \quad , \text{ wobei } W^{(i)} \text{ die Verteilung } K(z - \theta_o) \text{ hat, } i = 1, \ldots, N .$$

Beweis: Wegen $(\ln k(z-\phi))'(\theta_o) = -\frac{\partial}{\partial z} \ln k(z-\theta_o)$

folgt unmittelbar die Behauptung.

Beispiele:

1. logistische Verteilung:

$$K(z) = \frac{1}{1+e^{-z}} \qquad k(z) = \frac{e^{-z}}{(1+e^{-z})^2}$$

$$K(z-\phi) = \frac{1}{1+e^{-(z-\phi)}} \qquad k(z-\phi) = \frac{e^{-(z-\phi)}}{\left(1+e^{-(z-\phi)}\right)^2}$$

Es gilt:

$$-\frac{\partial}{\partial W^{(t_j)}} \ln k\left(W^{(t_j)}\right) = 2K\left(W^{(t_j)}\right) - 1 .$$

Da W die Verteilungsfunktion K hat, ist $H = K(W)$ gleichverteilt über $(0,1)$.

Es folgt:

$$-\frac{\partial}{\partial W^{(t_j)}} \ln k\left(W^{(t_j)}\right) = 2H^{(t_j)} - 1 ,$$

wobei die $H^{(i)}$ über $(0,1)$ gleichverteilt sind $(i = 1,\ldots, N)$.

Wegen $E\left[H^{(t_j)}\right] = \frac{t_j}{N+1}$ ist also der lokal beste Rangtest für Lagealternativen bezüglich einer logistischen Verteilung bestimmt durch :

$$U = \sum_{j=1}^{n} t_j \quad ,$$

der bekannten Mann-Whitney-Wilcoxon Statistik, häufig kurz U-Statistik genannt.

2. Normalverteilung

$$k(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} z^2}$$

$$k(z - \phi) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} (z - \phi)^2}$$

Es gilt :

$$-\frac{\partial}{\partial W^{(t_j)}} \ln k\left(W^{(t_j)}\right) = W^{(t_j)} \quad .$$

Somit ist der lokal beste Rangtest für Lagealternativen bezüglich einer Normalverteilung bestimmt durch:

$$C_1 = \sum_{j=1}^{n} E\left(W^{(t_j)}\right) \quad ,$$

wobei $W^{(i)}$ die standardisierte Normalverteilung hat (i = 1,...,N).

Die Teststatistik C_1 ist die bekannte Fischer-Yates-Statistik.

Korollar 2 (Variabilitätsalternativen)

Die Voraussetzungen des Satzes 8 seien erfüllt für die Verteilungsfunktionen

$$F_1(z) = K\left(\frac{z}{\theta}\right) \qquad \text{und} \qquad F_2(z) = K\left(\frac{z}{\phi}\right)$$

mit den Dichten

$$\frac{1}{\theta} k\left(\frac{z}{\theta}\right) \qquad \text{bzw.} \qquad \frac{1}{\phi} k\left(\frac{z}{\phi}\right) \quad , \theta, \phi \in \mathbb{R}^+ .$$

Der lokal beste Rangtest für

$$H_o : \theta = \phi =: \theta_o \quad \text{gegen}$$
$$H_1 : \phi > \theta \ (\phi < \theta)$$

wird beschrieben durch die Statistik V mit

$$V = - \sum_{j=1}^{n} E_{\theta_o \theta_o}\left[W^{(t_j)} \cdot \frac{\partial}{\partial W^{(t_j)}} \ln k\left(W^{(t_j)}\right)\right] \quad , \text{ wobei } W^{(i)}$$

die Verteilung $K\left(\frac{z}{\theta_o}\right)$ hat, $i = 1,\ldots,N$.

Beweis: Nach Satz 8 muß

$$E_{\theta_o \theta_o}\left[\left(\ln k\left(W^{(t_j)},\phi\right)\right)'(\theta_o)\right] \quad \text{bestimmt werden.}$$

Wegen $k(z,\phi) = \frac{1}{\phi} k\left(\frac{z}{\phi}\right)$ folgt:

$\ln k(z,\phi) = - \ln \phi + \ln k\left(\frac{z}{\phi}\right)$ und damit:

$(\ln k(z,\phi))' = - \frac{1}{\phi} - \frac{z}{\phi} \frac{\partial}{\partial z} \ln k\left(\frac{z}{\phi}\right)$.

Es gilt also

$$(\ln k(z,\phi))'(\theta_o) = -\frac{1}{\theta_o} - \frac{z}{\theta_o}\frac{\partial}{\partial z}\ln k\left(\frac{z}{\theta_o}\right).$$

Daraus folgt sofort die Behauptung.

Beispiele:

3. Normalverteilung

$$k(z) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}z^2}$$

$$\frac{1}{\phi}\, k\left(\frac{z}{\phi}\right) = \frac{1}{\sqrt{2\pi}\,\phi}\, e^{-\frac{1}{2}\left(\frac{z}{\phi}\right)^2} .$$

Es gilt:

$$-\frac{\partial}{\partial W^{(t_j)}} \ln k\left(W^{(t_j)}\right) = W^{(t_j)},$$

d.h. der lokal beste Rangtest für Variabilitätsalternativen bezüglich einer Normalverteilung ist bestimmt durch:

$$C = \sum_{j=1}^{n} E\left[\left(W^{(t_j)}\right)^2\right],$$

wobei $W^{(i)}$ die standardisierte Normalverteilung hat ($i = 1,\ldots,N$). C wird Capon-Statistik genannt.

4. Exponentialverteilung

$$k(z) = \begin{cases} e^{-z} & \text{für} \quad z > 0 \\ 0 & \text{sonst} \end{cases}$$

$$\frac{1}{\phi} k\left(\frac{z}{\phi}\right) = \begin{cases} \frac{1}{\phi} e^{-\frac{z}{\phi}} & \text{für} \quad z > 0 \\ 0 & \text{sonst} \end{cases}$$

Es ist:

$$-\frac{\partial}{\partial W^{(t_j)}} \ln k\left(W^{(t_j)}\right) = \begin{cases} 1 & \text{für} \quad z > 0 \\ 0 & \text{sonst} \end{cases}$$

Da gilt:[1)]

$$E\left(W^{(t_j)}\right) = \sum_{i=N-t_j+1}^{N} \frac{1}{i} \quad ,$$

ist also der lokal beste Rangtest für Variabilitätsalternativen bezüglich einer Exponentialverteilung bestimmt durch:

$$S = \sum_{j=1}^{n} \sum_{i=N-t_j+1}^{N} \frac{1}{i} \quad .$$

Anmerkung:

Moses zeigt an dem Beispiel[2)] zweier Dreiecksverteilungen, daß Rangtests auf Dispersion ungeeignet sind, wenn nicht starke Bedingungen an die beiden Verteilungsfunktionen F_1 und F_2 gestellt werden; so wie es in (4b) in 3.2 der Fall ist, daß F_1 und F_2 gleichen Median oder Mittelwert haben.

---

1) Gumbel, [22], S. 117

2) Moses, [43], S. 979

Werden z.B. F_1 und F_2 so gewählt, daß für ein $a \in \mathbb{R}$ gilt $F_1(a) = 0$ und $F_2(a) = 1$, dann kann kein Rangtest den Dispersionsunterschied "genügend" aufdecken. Die Überlegungen basieren auf der Tatsache, daß echt monoton wachsende Abbildungen die Verteilung der Ränge unverändert lassen, wohingegen die "Distanzen" verzerrt werden können.

4.3 Lokal beste Rangtests für eine Mischungsalternative[1)]

Für die in (5) in 3.2 angegebene spezielle Mischungsalternative

$$K(z,\phi) = (1 - \phi + \theta)\, K(z,\theta) + (\phi - \theta)\, K^2(z,\theta)$$

soll nun für $\phi > \theta$ (analog $\phi < \theta$) ein lokal bester Rangtest hergeleitet werden. Dazu werden zunächst die

$$c_{t_j} = E_{\theta_o \theta_o} \left[\ln k\left(W^{(t_j)}, \phi\right)'(\theta_o) \right]$$

aus Satz 8 anders dargestellt, was sich für manche Berechnungen des Erwartungswertes als günstiger erweist. Die Voraussetzungen von Satz 8 seien erfüllt. Es sei $f_{\theta\phi} : [0,1] \to [0,1]$ mit

$$f_{\theta\phi}\left[K(z,\theta)\right] = K(z,\phi) .$$

Da $k(z,\theta)$ und $k(z,\phi)$ existieren, gilt:

$$\frac{k(z,\phi)}{k(z,\theta)} = f^*_{\theta\phi}\left[K(z,\theta)\right]$$

mit $$f^*_{\theta\phi}(u) := \frac{\partial f_{\theta\phi}(u)}{\partial u} .$$

---

1) Pfanzagl, [46], S. 143

Dann folgt:

$$\frac{\partial}{\partial\phi} f^*_{\theta\phi}\left[K(z,\theta)\right]\Big|_{\theta=\phi=\theta_o} = \frac{\partial}{\partial\phi} \ln k(z,\phi)\Big|_{\phi=\theta_o}$$

$$= : (\ln k(z,\phi))'(\theta_o).$$

Nach Satz 8 gilt dann:

$$c_{t_j} = E_{\theta_o\theta_o}\left[\frac{\partial}{\partial\phi} f^*_{\theta\phi}\left[K\left(W^{(t_j)}, \theta\right)\right]\Big|_{\theta=\phi=\theta_o}\right].$$

Da $W^{(i)}$ nach Satz 8 die Verteilung K hat, ist $U^{(i)} := K(W^{(i)})$ über (0,1) gleichverteilt, d.h.

$$c_{t_j} = E\left[\frac{\partial}{\partial\phi} f^*_{\theta\phi}\left(U^{(t_j)}\right)\Big|_{\theta=\phi=\theta_o}\right].$$

Es sei nun $u = K(z,\theta)$ mit

$$K(z,\phi) = f_{\theta\phi}(u) = (1 - \phi + \theta)\, u + (\phi - \theta)\, u^2 ,$$

dann gilt:

$$\frac{\partial}{\partial\phi} f^*_{\theta\phi}(u)\Big|_{\theta=\phi=\theta_o} = -1 + 2u , \quad 0 \le u \le 1 .$$

Es folgt:

$c_{t_j} = -1 + 2\,E\left(U^{(t_j)}\right)$. Wegen $E\left(U^{(t_j)}\right) = \frac{t_j}{N+1}$ liefert also die Wilcoxon-Statistik $\sum_{j=1}^{n} t_j$ einen lokal besten Rangtest für die obige (einseitige) Mischungsalternative, so wie sie einen solchen Rangtest für einseitige Lagealternativen bei Annahme einer logisti-

schen Verteilung bestimmt.

Gibbons[1] hat für einige andere Mischungsalternativen lokal beste Rangtests hergeleitet. Sie sind aber wohl kaum von praktischem Interesse.

4.4 Gütevergleich des besten Rangtests mit dem C_1 - Test und dem U - Test bei Normalalternativen

Mit Satz 8 wird die Existenz eines besten Rangtests in einer "hinreichend kleinen" Umgebung der Nullhypothese ausgesprochen, wenn gewisse Voraussetzungen erfüllt sind. Es stellt sich sofort die Frage, wie klein $\varepsilon(N)$ im konkreten Fall gewählt werden muß. Da ε nicht nur eine Funktion von N, sondern auch von α und insbesondere der speziellen Verteilungsfunktion ist (Beispiele 1. - 4.), kann die gestellte Frage nicht allgemein beantwortet werden. Es gibt meines Wissens bislang auch keinerlei Untersuchungen zu diesem Problemkreis. So ist mit der Existenz einer $\varepsilon(N)$-Umgebung, in der ein Rangtest gleichmäßig bester ist, zunächst mehr ein beweistechnisches Prinzip als ein konkreter Aussagewert verbunden.

Die folgende Tabelle A zeigt für Normalalternativen $d = \theta_Y - \theta_X$[2] mit $d = .2(.2)1{,}0;\ 1{,}5;\ 2{,}0;\ 3{,}0$ und für bestimmte Kombinationen (m,n) der Stichprobenumfänge von $x_1, \ldots, x_m$, $y_1, \ldots, y_n$ einen Gütevergleich zwischen dem besten Rangtest B, dem C_1 - Test als lokal bestem Rangtest und dem U - Test. Das Niveau α ist hier durch $\alpha = \frac{q}{\binom{n+m}{n}}$ festgelegt mit $\alpha \leq 0{,}1$; d.h. genau q Elemente

---

1) Gibbons, [21], S. 293

2) θ_X und θ_Y sind die Erwartungswerte der normalverteilten Stichprobenvariablen X_i bzw. Y_j, $i=1,\ldots,m$, $j=1,\ldots,n$.

(Rangkombinationen) fallen in den kritischen Bereich (Seite 43).

Der Gütevergleich erfolgt mit Hilfe eines von R.C.Milton erstellten Tabellenwerks über "Rank order probabilities", das im Jahre 1970 erschienen ist und eine große Lücke auf dem Gebiet der Rangtests für Normalalternativen der Lage im Zwei-Stichproben-Problem geschlossen hat.

Für die Kombinationen (m,n) mit $m + n \leq 8$, $\alpha \leq 0{,}1$ stimmen die Gütewerte des B - Tests, C_1 - Tests und des U - Tests bis auf einige wenige Ausnahmefälle α und d überein.

Tabelle A verdeutlicht zunächst, welchen Einfluß α und d auf die Güten von B, U und C_1 haben. Der C_1 - Test ist für $d = 0{,}2$ bei fast allen Stichprobenkombinationen bester Rangtest oder zumindest besser als der U - Test, seine optimale lokale Eigenschaft wird also bezüglich der hier gewählten (m,n) - α - Kombinationen durch diese Übersicht im wesentlichen bestätigt. Eine erwähnenswerte Ausnahme bildet das Beispiel (7,6) mit $\alpha \approx 0{,}01$, in dem der U - Test sogar für alle d bester Rangtest ist. Für $d = 2{,}0$ und $d = 3{,}0$ fällt darüber hinaus der Gütevergleich zwischen dem U - Test und dem C_1 - Test in den meisten Fällen zugunsten des U - Tests aus. Diese oftmals höhere "Effizienz" des U - Tests gegenüber dem C_1 - Test für größere Distanzen von der Nullhypothese findet mit den Ausführungen des folgenden Kapitels ihre Bestätigung.

Tabelle A

(m,n)	d	α	Gütevergleich
(5,4)	0,2 - 0,6	12/126 ≃ 0,095	$B = C_1 > U$
	0,8 - 3,0		$B = U > C_1$
(6,3)	0,2 - 2,0	7/84 ≃ 0,083	$B = C_1 > U$
	3,0		$B = U > C_1$
(5,5)	0,2 - 1,0	12/252 ≃ 0,048	$B = C_1 > U$
	1,5 - 3,0		$B = U > C_1$
	0,2 - 3,0	16/252 ≃ 0,064	$B = C_1 > U$
(6,4)	0,2 - 1,0	12/210 ≃ 0,057	$B = C_1 > U$
	1,5		$B > U > C_1$
	2,0 - 3,0		$B = U > C_1$
	0,2 - 3,0	18/210 ≃ 0,086	$B = C_1 > U$
(7,3)	0,2 - 3,0	7/120 ≃ 0,058	$B = C_1 > U$
	0,2 - 3,0	11/120 ≃ 0,092	$B = C_1 > U$
(6,5)	0,2 - 3,0	12/462 ≃ 0,026	$B = C_1 = U$
	0,2 - 1,0	29/462 ≃ 0,063	$B = C_1 > U$
	1,5		$B > C_1 > U$
	2,0 ; 3,0		$B > U > C_1$
(7,4)	0,2 - 1,5	18/330 ≃ 0,055	$B = C_1 > U$
	2,0 ; 3,0		$B > C_1 > U$
	0,2 - 2,0	27/330 ≃ 0,082	$B = C_1 > U$
	3,0		$B > C_1 > U$
(6,6)	0,2 - 1,5	18/924 ≃ 0,019	$B = C_1 > U$
	2,0 ; 3,0		$B = U > C_1$
(7,5)	0,2 - 2,0	19/792 ≃ 0,024	$B = C_1 > U$
	3,0		$B > U > C_1$
	0,2 - 1,0	29/792 ≃ 0,037	$B > C_1 > U$
	1,5 - 3,0		$B > U > C_1$
(7,6)	0,2 - 3,0	19/1716 ≃ 0,011	$B = U > C_1$
	0,2	63/1716 ≃ 0,036	$B = C_1 > U$
	0,4 - 1,5		$B > C_1 > U$
	2,0 ; 3,0		$B > U > C_1$
(7,7)	0,2	18/3432 ≃ 0,005	$B = C_1 > U$
	0,4 - 3,0		$B = U > C_1$
	0,2 - 1,0	37/3432 ≃ 0,011	$B > C_1 > U$
	1,5 - 3,0		$B > U > C_1$

5. Kapitel Relative Effizienz ausgesuchter Rangtests für das Zwei-Stichproben-Problem

5.1 Definitionen der relativen Effizienz

Im Rahmen einer nichtparametrischen Testtheorie bedeutet die Suche nach optimalen Tests allein unter den Rangtests eine nennenswerte Einschränkung. Um über die Bevorzugung eines besten oder lokal besten Rangtests entscheiden zu können, müßten diese noch hinsichtlich ihrer Güteeigenschaften mit anderen nichtparametrischen Verfahren, wie den Tests vom K.-S.-Typ oder den Permutationstests verglichen werden. Noch wichtiger aber scheint mir die Fragestellung zu sein, ob Rangtests in ihrer "Wirksamkeit" einem Vergleich mit den entsprechenden parametrischen Verfahren standhalten, so z.B. mit dem t- Test und F- Test im Zwei-Stichproben-Problem bei Tests auf Lage bzw. Variabilität, auch unter der Annahme nichtnormalverteilter Grundgesamtheiten. Gerade die Beantwortung der Frage, inwieweit Rangtests bei Annahme einer bestimmten Verteilungsfunktion und ihrer Parameter in Konkurrenz mit ihren verteilungsgebundenen "Gegenstücken" treten können, ja sogar wirksamer als diese sein können, ist von besonderem Interesse, wird doch damit sozusagen eine "Rückversicherung" gewährleistet.

Eine Möglichkeit des Gütevergleichs zweier Tests besteht in der Bestimmung der sogenannten relativen Effizienz.

Es seien $[\varphi_n^{(1)}]$ und $[\varphi_n^{(2)}]$ zwei Testfolgen[1] vom Niveau α für dieselbe Null- und Alternativhypothese und $[\beta_n^{(1)}]$ bzw. $[\beta_n^{(2)}]$ die zugehörigen Folgen der Gütefunktionen.

---

1) d.h. für jedes $n \geq 1$ über dem Stichprobenraum definierte $\mathcal{A}$-meßbare Funktionen $\varphi_n^{(i)}: S \to [0,1]$; $i=1,2$

Ferner gelte:

$$\bigwedge_{n \in \mathbb{N}} \quad \bigwedge_{\theta \in \Omega - \omega} \quad \beta_n^{(2)}(\theta) > \beta_n^{(1)}(\theta) \ .$$

Jeder Test $\varphi_n^{(1)}$ der Folge $[\varphi_n^{(1)}]$ wird jedoch dem Test $\varphi_n^{(2)}$ von $[\varphi_n^{(2)}]$ vorgezogen, weil er z.B. geringeren Rechenaufwand erfordert oder größere Anwendungsgültigkeit hat (keine Annahme einer bestimmten Verteilungsfunktion, wie die der Normalverteilung, keine Forderung einer zugrunde liegenden Intervallskala o.ä.).

Es stellt sich folgende Frage:

Sollen die Kosten der Untersuchung (als monoton wachsende Funktion des Stichprobenumfanges n) möglichst niedrig gehalten und damit ein gewisser Güteverlust bei Anwendung von $\varphi_n^{(1)}$ statt $\varphi_n^{(2)}$ in Kauf genommen werden oder können die Kosten eine untergeordnete Rolle spielen, wenn nur kein Güteverlust bei Anwendung von $\varphi_n^{(1)}$ statt von $\varphi_n^{(2)}$ eintritt?

Drei Verlustfunktionen müssen also berücksichtigt werden:

(1) Güteverlust
(2) Kosten der Untersuchung
(3) Nachteile bei Anwendung von $\varphi_n^{(2)}$.

In der Praxis ist natürlich (3) kaum abzuschätzen und nicht in einen direkten Zusammenhang mit (1) und (2) zu bringen. Es ergeben sich somit 2 spezielle Probleme:

I Gütevergleich bei festem Stichprobenumfang n
Als Maße des Güteverlustes können die Funktionen

V_i: (0,1) x (0,1) → (0,1) , i = 1, 2, mit :

$$V_1(\beta_n^{(1)}, \beta_n^{(2)}) = \beta_n^{(2)}(\theta) - \beta_n^{(1)}(\theta)$$

oder

$$V_2(\beta_n^{(1)}, \beta_n^{(2)}) = 1 - \frac{\beta_n^{(1)}(\theta)}{\beta_n^{(2)}(\theta)}$$

gewählt werden.

II Vergleich der beiden Stichprobenumfänge n_1 von $\varphi_{n_1}^{(1)}$ und n_2 von $\varphi_{n_2}^{(2)}$ bei fest vorgegebener Güte $\beta(\theta)$, $\theta \in \Omega - \omega$.

Es sei:

(1) $$n_1(\beta_1) := \min \{n' \in \mathbb{N} \mid \beta_{n'}^{(1)} \geq \beta(\theta) \}$$

$$n_2(\beta) := \min \{n'' \in \mathbb{N} \mid \beta_{n''}^{(2)} \geq \beta(\theta) \} .$$

Als Maße des Verlustes können die Funktionen K_i: $\mathbb{N} \times \mathbb{N} \to \mathbb{R}^+$, i = 1,2 mit:

$$K_1(n_1, n_2) = k(n_1) - k(n_2)$$ oder

$$K_2(n_1, n_2) = 1 - \frac{k(n_2)}{k(n_1)}$$ gewählt werden, wobei

die Funktion $k: \mathbb{N} \to \mathbb{R}^+$ die Kosten der Untersuchung in Abhängigkeit vom Stichprobenumfang angibt.

Die Definition einer relativen Effizienz bezüglich zweier Testverfahren ist sowohl nach I als auch nach II möglich. So kann in I der Quotient $\frac{\beta_n^{(1)}(\theta)}{\beta_n^{(2)}(\theta)}$, $n \in \mathbb{N}$, $\theta \in \Omega - \omega$ fest, als (finite) relative Effizienz gekennzeichnet und unmittelbar aus den entsprechenden Tabellen der Gütewerte berechnet werden.

Mir ist jedoch keine Publikation bekannt geworden, in der Effizienzen von Rangtests bezüglich der entsprechenden parametrischen Verfahren für das Zwei-Stichproben-Problem über

$\frac{\beta_n^{(1)}(\theta)}{\beta_n^{(2)}(\theta)}$ angegeben sind.

Die herkömmliche Definition der relativen Effizienz zweier Testverfahren bezieht sich auf die in II speziell zu wählende Funktion $K: \mathbb{N} \times \mathbb{N} \to (0,1)$ mit $K(n_1, n_2) = \frac{n_2}{n_1}$. Darin wird der Quotient $\frac{n_2}{n_1}$ allgemein als relative Effizienz von $\varphi_n^{(1)}$ bezüglich $\varphi_n^{(2)}$ bezeichnet; für $n_1, n_2 \in \mathbb{N}$ als finite relative Effizienz (F.R.E.) und für $n_1, n_2 \to \infty$ als asymptotische relative Effizienz (A.R.E.). Die Frage nach der Existenz des Grenzwertes von $\frac{n_2}{n_1}$ wird im nächsten Abschnitt beantwortet.

5.2 Asymptotische relative Effizienz unter besonderer Berücksichtigung des Pitmanschen Konzepts

Die Existenz von $\lim_{n_1, n_2 \to \infty} \frac{n_2}{n_1}$ ist natürlich nur unter ganz bestimmten Voraussetzungen gesichert, die im Zusammenhang mit 3 möglichen A.R.E.-Definitionen stehen. Diese 3 Fälle seien im folgenden bezüglich des Testproblems $(\alpha, \omega, \Omega - \omega)$ schematisch dargestellt:

(A) α fest, β als asymptotische Güte fest vorgegeben; Vergleich von Folgen $[\theta_n]$, $\theta_n \in \Omega - \omega$ mit $\theta_n \to \theta_o$ und $\omega = \{ \theta_o \}$ (Pitman)

(B) α fest, $\theta \in \Omega - \omega$ fest, Vergleich von Folgen $[\beta_n]$ an der Stelle θ (Hodges/Lehmann)[1)]

(C) β fest, $\theta \in \Omega - \omega$ fest, Vergleich von Folgen $[\alpha_n] \to 0$ (Bahadur)[2)]

Die Definition der A.R.E. bezüglich zweier Tests ist also völlig willkürlich. Ihre Fragwürdigkeit wird durch eine Untersuchung von Tsutakawa[3)] unterstrichen, der an einem Beispiel zeigt, daß für 2 Testfolgen die nach (A), (B) und (C) berechneten A.R.E.'s sehr verschieden voneinander sein können.

Wenn auch in jüngster Zeit das Konzept von Bahadur immer mehr an Bedeutung gewinnt, so ist doch die Definition von Pitman bislang am meisten angewandt worden; nicht zuletzt deshalb, weil es auch gerade von der Praxis her gerechtfertigt erscheint, Alternativen in der Nähe

---

1) Hodges/Lehmann, [27], S. 324

2) Bahadur, [1], S. 276

3) Tsutakawa, [62], S. 179

der Nullhypothese zu betrachten. So kann die Pitmansche A.R.E. auch als lokale A.R.E.bezeichnet werden. Da der noch folgenden Zusammenstellung der A.R.E.'s einiger Rangtests bezüglich des t-Tests bzw. F-Tests die Definition von Pitman zugrunde liegt, sei darauf zunächst näher eingegangen.

Def. 14 : Eine Folge von Tests $[\varphi_n]$ für das Problem $(\alpha, \omega, \Omega - \omega)$ heißt konsistent, wenn gilt:

$$\bigwedge_{\theta \in \Omega - \omega} \lim_{n \to \infty} \beta_n(\theta) = 1$$

Da die meisten Testfolgen - so auch die in diesem Kapitel betrachteten Rangtests - diese "Minimalforderung" der Konsistenz erfüllen, kann also der Limes der Gütefunktionen bei festem $\theta \in \Omega - \omega$ im allgemeinen nicht als Vergleichskriterium zweier Testfolgen herangezogen werden. Ein sinnvoller asymptotischer Vergleich zweier konsistenter Testfolgen $[\varphi_n^{(1)}]$ und $[\varphi_n^{(2)}]$ ist jedoch häufig dadurch möglich, daß die Folgen der Gütefunktionen $[\beta_n^{(1)}]$ und $[\beta_n^{(2)}]$ von $[\varphi_n^{(1)}]$ bzw. $[\varphi_n^{(2)}]$ in Abhängigkeit von $[\theta_n]$, $\theta_n \in \Omega - \omega$, $\theta_n \to \theta_o$ untersucht werden. Das ist das Konzept der Pitmanschen A.R.E.:

Def. 15 : Es seien $[\varphi_n^{(1)}]$ und $[\varphi_n^{(2)}]$ zwei Testfolgen für das Problem $(\alpha, \theta_o, \Omega - \{\theta_o\})$; $[\beta_n^{(1)}(\theta_n)]$ bzw. $[\beta_n^{(2)}(\theta_n)]$ seien die zugehörigen Folgen der Gütefunktionen für eine Folge $[\theta_n]$, $\theta_n \in \Omega - \{\theta_o\}$ mit $\lim_{n \to \infty} \theta_n = \theta_o$.

Weiterhin seien $n_1^{(i)}$ und $n_2^{(i)}$ zwei monoton wachsende Folgen natür-

licher Zahlen, für die gilt:

$$\lim_{n_1^{(i)}\to\infty} \beta^{(1)}_{n_1^{(i)}}(\theta_i) = \lim_{n_2^{(i)}\to\infty} \beta^{(2)}_{n_2^{(i)}}(\theta_i) = \beta \ , \quad 0 < \beta < 1 \ .$$

Dann ist die A.R.E. von $[\varphi_n^{(1)}]$ bezüglich $[\varphi_n^{(2)}]$ nach Pitman definiert durch :

$$\text{Eff.}\left([\varphi_n^{(1)}], [\varphi_n^{(2)}], \alpha, \theta_i\right) : = \lim_{i\to\infty} \frac{n_2^{(i)}}{n_1^{(i)}} \ ,$$

vorausgesetzt, daß dieser Limes existiert und für jede Wahl von $n_1^{(i)}$, $n_2^{(i)}$ derselbe ist.

Der folgende Satz gibt an, wann der Grenzwert existiert und wie er dann numerisch bestimmt werden kann.

Satz 9 :[1] Es sei $[\varphi_n]$ eine Testfolge für $H_o : \theta = \theta_o$ gegen $H_1 : \theta = \theta_n = \theta_o + \frac{k}{\sqrt{n}}$, $k \in \mathbb{R}^+$ [2], die durch die Folge der Teststatistiken $[t_n]$ beschrieben werde. Es bezeichne:

$$E_\theta(t_n) = \mu_n(\theta) \ , \quad \text{Var}_\theta(t_n) = \sigma_n^2(\theta) \ .$$

Folgende Bedingungen seien erfüllt:

(i) $\lim_{n\to\infty} P_{\theta_o}(t_n \geq c_{n,\alpha}) = \alpha, \ 0 < \alpha < 1$,

(ii) Die Verteilung von t_n ist stetig, und $[t_n]$ ist asymptotisch normalverteilt unter H_o und H_1 , und zwar gleichmäßig in

---

1) Noether, [44], S. 64

2) Der Fall $k \in \mathbb{R}^-$, d.h. also $\theta < \theta_o$, kann entsprechend behandelt werden.

θ mit $\theta_o \leq \theta \leq \theta_o + a$, $a > 0$,

(iii) $\mu_n'(\theta)$: $= \frac{\partial \mu_n(\theta)}{\partial \theta}$ existiert für alle $\theta \in [\theta_o, \theta_o + a]$,

und es gelte : $\mu_n'(\theta_o) \neq 0$,

(iv) $\lim\limits_{n\to\infty} \frac{\mu_n'(\theta_n)}{\mu_n'(\theta_o)} = 1$ und $\lim\limits_{n\to\infty} \frac{\sigma_n(\theta_n)}{\sigma_n(\theta_o)} = 1$,

(v) $\lim\limits_{n\to\infty} \frac{1}{\sqrt{n}} \frac{\mu_n'(\theta_o)}{\sigma_n(\theta_o)}$ existiert und sei gleich s , $s \in \mathbb{R}$.

Dann gilt für die asymptotische Güte von $[\varphi_n]$:

$$\lim_{n\to\infty} P_{\theta_n}(t_n \geq c_{n,\alpha}) = 1 - \Phi(c_\alpha - ks) ,$$

wobei $1 - \Phi(c_\alpha) = \alpha$ ist und Φ die standardisierte Normalverteilung.

Beweis: Wegen (ii) gilt:

$$\lim_{n\to\infty} P_{\theta_n}(t_n \geq c_{n,\alpha}) = 1 - \lim_{n\to\infty} \Phi\left(\frac{c_{n,\alpha} - \mu_n(\theta_n)}{\sigma_n(\theta_n)}\right)$$

$$= 1 - \Phi(w) \quad \text{mit} \quad w = \lim_{n\to\infty} \frac{c_{n,\alpha} - \mu_n(\theta_n)}{\sigma_n(\theta_n)} .$$

Aus $\mu_n(\theta_n) = \mu_n(\theta_o + \frac{k}{\sqrt{n}}) = \mu_n(\theta_o) + \frac{k}{\sqrt{n}} \mu_n'(\tilde{\theta})$ mit $\theta_o < \tilde{\theta} < \theta_n$

folgt:

$$w = \lim_{n\to\infty} \frac{c_{n,\alpha} - \mu_n(\theta_o) - \frac{k}{\sqrt{n}}\,\mu_n'(\tilde{\theta})}{\sigma_n(\theta_n)} =$$

$$= \lim_{n\to\infty} \left[\frac{c_{n,\alpha} - \mu_n(\theta_o) - \frac{k}{\sqrt{n}}\,\mu_n'(\tilde{\theta})}{\sigma_n(\theta_n)} \cdot \frac{\sigma_n(\theta_o)}{\sigma_n(\theta_o)} \right] .$$

Wegen $$\lim_{n\to\infty} \frac{c_{n,\alpha} - \mu_n(\theta_o)}{\sigma_n(\theta_o)} = c_\alpha$$

und (iv), (v) gilt dann:

$$w = c_\alpha - ks \,. \qquad \text{q.e.d.}$$

Korollar: Sind $[\varphi_n^{(1)}]$ und $[\varphi_n^{(2)}]$ zwei Testfolgen vom Niveau α und $[t_n^{(1)}]$ bzw. $[t_n^{(2)}]$ die Folgen der zugehörigen Teststatistiken, die die Voraussetzungen des obigen Satzes erfüllen mit

$$E_\theta\left(t_n^{(i)}\right) = \mu_{in}(\theta) \,, \quad \mathrm{Var}_\theta\left(t_n^{(i)}\right) = \sigma_{in}^2(\theta)$$

und $$\lim_{n\to\infty} \frac{1}{\sqrt{n}} \frac{\mu_{in}'(\theta_o)}{\sigma_{in}(\theta_o)} = s_i \,, \quad i = 1,2$$

dann gilt:

$$\mathrm{Eff.}\left([\varphi_n^{(1)}],[\varphi_n^{(2)}]\right) = \lim_{n\to\infty} \frac{n_2^{(i)}}{n_1^{(i)}} = \frac{s_1^2}{s_2^2}$$

$$= \lim_{n\to\infty} \left(\frac{\mu_{1n}'(\theta_o)}{\mu_{2n}'(\theta_o)} \cdot \frac{\sigma_{2n}(\theta_o)}{\sigma_{1n}(\theta_o)} \right)^2 .$$

Beweis: Sollen die beiden Testfolgen dieselbe asymptotische Güte gegen dieselbe Folge der Alternativen für festes α haben (Definition 15), so muß nach obigem Satz gelten:

$$k_1 s_1 = k_2 s_2 \qquad \text{und}$$

$$\frac{k_1}{\sqrt{n_1^{(i)}}} = \frac{k_2}{\sqrt{n_2^{(i)}}} \quad .$$

Daraus folgt:

$$\lim_{i \to \infty} \frac{n_2^{(i)}}{n_1^{(i)}} = \frac{s_1^2}{s_2^2} \quad . \qquad \text{q.e.d.}$$

Anmerkung:

Die Forderung (ii) in Satz 9 bedeutet keine starke Einschränkung, denn die meisten Teststatistiken haben diese Eigenschaft.[1] Sie kann jedoch durch die schwächere Bedingung ersetzt werden, daß die Verteilungen von $\left[t_n^{(1)} \right]$ und $\left[t_n^{(2)} \right]$ unter H_o und H_1 gegen die gleiche (aber beliebige) stetige, echt monotone Verteilungsfunktion konvergieren. Dadurch ist jedenfalls ein (asymptotischer) Gütevergleich möglich.

---

1) Der allgemeine Beweis der asymptotischen Normalität für Rangstatistiken der Form $t = \sum h(r_i)$, die gewissen Bedingungen genügen, ist von Chernoff/Savage, [11], S. 972 erbracht worden.

5.3 Die Pitmansche A.R.E. einiger Rangtests zum t-Test bzw. F-Test

Die folgende Übersicht zeigt einige ausgewählte Rangstatistiken für Tests auf Lage und Variabilität (unter ihnen die bereits im letzten Kapitel hergeleiteten Tests von Wilcoxon, Fischer-Yates und Capon) mit einer A.R.E.-Tabelle B für verschiedene Verteilungen der Grundgesamtheit, von denen gewisse mehr von theoretischem als von praktischem Interesse sein mögen. Einige Werte habe ich selbst berechnet, die anderen der Literatur entnommen, auf die im einzelnen verwiesen wird und in der auch der Beweis für das Erfülltsein der Bedingungen (iii) - (v) bei Anwendung der A.R.E.-Formel nach Pitman geführt ist.

Die angegebenen Werte sprechen für eine bevorzugte Anwendung der Rangtests auf Lage gegenüber dem t-Test. So fällt die A.R.E. des C_1-Tests und des X-Tests (die beiden Tests heißen asymptotisch äquivalent[1]) zum t-Test nicht unter 1, und dieser Wert wird nur für die Normalverteilung angenommen.[2] Auch die untere Grenze der A.R.E. des U-Tests zum t-Test ist mit 0,864 sehr hoch. Da der U-Test zudem kaum Rechenaufwand erfordert und für ihn ein weit umfangreicheres Tabellenwerk vorliegt als für den C_1-Test und den X-Test, sollte er in der Praxis diesen Tests vorgezogen werden; besonders dann, wenn keine Normalverteilung angenommen wird, sondern Verteilungen mit "heavy tails" (logistische, doppelexponentiale u.a.).

Da der F-Test weniger robust als der t-Test gegenüber Abweichungen von der Normalverteilung ist[3], ist es nicht sinnvoll, A.R.E.'s für die angegebenen Rangtests auf Variabilität bezüglich des F-Tests bei nichtnormalverteilten Grundgesamtheiten zu berechnen. In diesem Fall sollte stets einer der Rangtests dem F-Test vorgezogen werden, wenn auch die untere Grenze ihrer A.R.E.'s gleich 0 ist. Da der K-Test leichter zu tabellieren ist als der C-Test, verdient er in der Praxis bevorzugte Anwendung.

---

1) Wittig, [69], S. 163
2) Gastwirth/Wolff, [19], S. 2188
3) Lehmann, [36], S. 173

Tabelle B

Teststatistiken

t_j sind die (geordneten) Ränge der y_j in der kombinierten Stichprobe $x_1, \ldots, x_m, y_1, \ldots, y_n$; Φ ist die standardisierte Normalverteilung

I Lage	Wilcoxon U-Test	$U = \sum_{j=1}^{n} t_j$	Fischer-Yates C_1-Test	$C_1 = \sum_{j=1}^{n} E\left[W^{(t_j)}\right]$	v.d.Waerden X-Test	$X = \sum_{j=1}^{n} \Phi^{-1}\left(\frac{t_j}{m+n+1}\right)$
II Variabilität	Mood M-Test	$M = \sum_{j=1}^{n} \left(t_j - \frac{m+n+1}{2}\right)^2$	Capon C-Test	$C = \sum_{j=1}^{n} E\left[\left(W^{(t_j)}\right)^2\right]$	Klotz K-Test	$K = \sum_{j=1}^{n} \left(\Phi^{-1}\left(\frac{t_j}{m+n+1}\right)\right)^2$

A.R.E. (nach Pitman)

	Vergleichender Test	Verteilung der Grundgesamtheit: normal	exponential	rechteckig	doppelexponential	logistisch	untere Grenze	obere Grenze
U -Test	t	0,955 [1]	3,000	1,000	1,500	1,096	0,864 [1]	∞ [1]
C_1-Test	t	1,000 [2]	∞	∞	1,273	1,047	1,000 [2]	∞
X -Test	t	1,000 [3]	∞	∞	1,273	1,047	1,000 [3]	∞
U -Test	C_1	0,955	0 [3]	0 [3]	1,178 [3]	1,047 [3]	0	1,910 [3]
M -Test	F	0,760 [4]					0 [5]	∞ [5]
C -Test	F	1,000 [6]					0 [7]	∞ [6]
K -Test	F	1,000 [6]					0 [7]	∞ [6]
M -Test	K	0,760 [6]	0,783 [8]	0 [6]	0,900 [6]	0,896 [6]		

1) Hodges/Lehmann, [27], S. 324
2) Chernoff/Savage, [11], S. 972
3) Hodges/Lehmann, [28], S. 307
4) Mood, [41], S. 514
5) Sukhatme, [61], S. 188
6) Klotz, [32], S. 498
7) Raghavachari, [50], S. 1306
8) BASU/Woodworth, [2], S. 276

Es stellt sich jedoch die Frage, welchen Aussagewert überhaupt die Pitmansche A.R.E. als eine einzige Zahl hat. Einerseits wird die Willkür ihrer Definition durch das erwähnte Beispiel von Tsutakawa unterstrichen, andererseits erscheint auch vom praktischen Standpunkt ein solches Vergleichskriterium als wenig sinnvoll ($n \to \infty$). Zudem liegen bislang keinerlei Untersuchungen darüber vor, wie groß n gewählt werden muß, damit z.B. der C_1-Test für nicht-normale Alternativen dem t-Test "überlegen" ist.

Somit ist es sicherlich realistischer, finite relative Effizienzen zu berechnen, in der Hoffnung, daß die so gewonnenen Ergebnisse im wesentlichen mit den A.R.E.'s, den "Orientierungszahlen", übereinstimmen. Hinzu kommt, daß Rangtests gerade für den Fall kleiner Stichproben Anwendung finden, ihre F.R.E.'s zum t-Test bzw. F-Test also von besonderem Interesse sind.

5.4 Finite relative Effizienzen des U-Tests und des C_1-Tests zum t-Test für Normalalternativen

5.4.1 F.R.E. für $m = n$ und einseitige Alternativen

Der zu Beginn dieses Kapitels über (1) (Seite 70) vorgeschlagene Ansatz zur Definition einer F.R.E. ist für kleine Stichprobenumfänge wenig geeignet und soll daher ersetzt werden durch:

$$(2) \qquad \beta^{(1)}(n_1, \alpha, \theta) = \beta^{(2)}(n_2, \alpha, \theta) ,$$

wobei wieder $\beta^{(i)}$ die zu $\varphi^{(i)}$ gehörige Gütefunktion $(i=1,2)$ und $\alpha, \theta \in H_1$ und z.B. $n_1 \in \mathbb{N}$ fest vorgegeben sind. In der Regel existiert dann keine natürliche Zahl n_2 , für die in (2) das Gleichheitszei-

chen gilt, wohl aber ein $\bar{n}_2 \in \mathbb{N}$ mit

$$\beta^{(2)}(\bar{n}_2,\alpha,\theta) \leq \beta^{(1)}(n_1,\alpha,\theta) \leq \beta^{(2)}(\bar{n}_2+1,\alpha,\theta) \quad .$$

Das führt dann zu einer "modifizierten" Definition der F.R.E., die für die folgenden Ausführungen herangezogen wird.

Def. 16 : Die F.R.E. von $\varphi^{(1)}$ bezüglich $\varphi^{(2)}$ ist erklärt durch $\frac{\tilde{n}_2}{n_1}$, $\tilde{n}_2 \in \mathbb{R}^+$, wobei $\tilde{n}_2$ zur Erfüllung der Gleichung (2) approximativ durch lineare Interpolation zwischen $\bar{n}_2$ und $\bar{n}_2 + 1$ bestimmt[1] wird, d.h.

$$\tilde{n}_2 = \bar{n}_2 + p \quad \text{mit} \quad p = \frac{\beta^{(1)}(n_1,\alpha,\theta) - \beta^{(2)}(\bar{n}_2,\alpha,\theta)}{\beta^{(2)}(\bar{n}_2+1,\alpha,\theta) - \beta^{(2)}(\bar{n}_2,\alpha,\theta)} \quad .$$

Dabei läßt sich $\tilde{n}_2$ wie folgt interpretieren:

Fassen wir den Umfang der Stichprobe bezüglich $\varphi^{(2)}$, für den die Gleichung (2) gilt, als Zufallsvariable N auf mit

$$P(N = \bar{n}_2) = 1 - p \qquad \text{und}$$
$$P(N = \bar{n}_2 + 1) = p \quad , \qquad \text{so ist}$$
$$E(N) = \bar{n}_2 + p = \tilde{n}_2 \quad .$$

Mit Hilfe der bereits erwähnten Tabellen von R.C.Milton über Gütewerte des U-Tests, C_1-Tests und des t-Tests bei Normalalternativen sind zunächst in Tabelle C über lineare Interpolation F.R.E.'s der beiden genannten Rangtests zum t-Test zusammengestellt, und zwar für:

---

1) Ein anderes Interpolationsverfahren zur Definition von $\tilde{n}_2$ wird im Abschnitt 5.4.4 angegeben.

a) einseitige Alternativen $\theta = d = \theta_Y - \theta_X$ mit

$d = .2(.2)1{,}0;\ 1{,}5;\ 2{,}0$ und $3{,}0$,

b) $m = n = 3,\ 4,\ 5,\ 6,\ 7$,

c) $\alpha = 0{,}01;\ 0{,}05$ und $0{,}1$.

In der Literatur sind solche F.R.E.'s bislang nur vereinzelt tabelliert worden, für den C_1-Test noch weniger als für den U-Test. So kann die oftmals geäußerte Vermutung, daß die F.R.E.'s des U-Tests zum t-Test sehr hoch sind und ungefähr um 0,95 liegen, wohl nur auf dem asymptotischen Wert (0,955) basieren.

Im zweiten Teil der Tabelle C sind für $m = n = 5,\ 6,\ 7$ und $\alpha = 0{,}01$ vergleichsweise finite relative Effizienzen über den Quotienten $\frac{\beta_n^{(1)}}{\beta_n^{(2)}}$ (nach den Ausführungen unter I in 5.1) aufgeführt.

Es zeigt sich, daß die so definierten F.R.E.'s im Unterschied zu den obigen mit wachsendem d schneller abnehmen und wieder stark ansteigen. Wenn sich auch $\frac{\beta_n^{(1)}}{\beta_n^{(2)}}$ unmittelbar (d.h. ohne Interpolation) berechnen läßt, so wird im folgenden nur noch auf die F.R.E.-Definition 16 zurückgegriffen, da die so bestimmten F.R.E.'s einen Vergleich mit der Pitmanschen A.R.E. ermöglichen, die auch über den Quotienten zweier Stichprobenumfänge definiert ist.

Anmerkung:
Da der t-Test gleichmäßig bester unverfälschter Test für einseitige Normalalternativen [1] ist und der U-Test und der C_1-Test unverfälscht sind für alle einseitigen Alternativen [2], müssen also die F.R.E.'s dieser beiden Rangtests zum t-Test für Normalalternativen kleiner als 1 sein.

---

1) Lehmann, [36], S. 172

2) Das ist eine unmittelbare Folgerung aus Satz 4.4, Schmetterer, [58], S. 540

Tabelle C

F.R.E. zum t-Test bei Normalalternativen für gleiche Stichprobenumfänge (einseitige Alternativen)

	m=n	α	α=0,2	0,4	0,6	0,8	1,0	1,5	2,0	3,0
U -Test	3	0,05	0,9818	0,9801	0,9784	0,9769	0,9755	0,9731	0,9724	0,9773
C_1-Test	3	0,05	0,9818	0,9801	0,9784	0,9769	0,9755	0,9731	0,9724	0,9773
U -Test	3	0,10	0,9787	0,9771	0,9756	0,9743	0,9731	0,9715	0,9719	0,9791
C_1-Test	3	0,10	0,9787	0,9771	0,9756	0,9743	0,9731	0,9715	0,9719	0,9791
U -Test	4	0,10	0,9582	0,9561	0,9541	0,9526	0,9514	0,9499	0,9508	0,9593
C_1-Test	4	0,10	0,9582	0,9561	0,9541	0,9526	0,9514	0,9499	0,9508	0,9593
U -Test	5	0,01	0,9645	0,9612	0,9580	0,9549	0,9520	0,9458	0,9415	0,9404
C_1-Test	5	0,01	0,9645	0,9612	0,9580	0,9549	0,9520	0,9458	0,9415	0,9404
U -Test	5	0,05	0,9597	0,9576	0,9559	0,9543	0,9531	0,9513	0,9515	0,9573
C_1-Test	5	0,05	0,9630	0,9609	0,9589	0,9572	0,9558	0,9535	0,9532	0,9580
U -Test	6	0,01	0,9667	0,9640	0,9613	0,9588	0,9564	0,9513	0,9476	0,9443
C_1-Test	6	0,01	0,9687	0,9657	0,9632	0,9604	0,9578	0,9519	0,9471	0,9400
U -Test	6	0,05	0,9506	0,9449	0,9447	0,9446	0,9440	0,9440	0,9451	0,9516
C_1-Test	6	0,05	0,9595	0,9576	0,9559	0,9544	0,9531	0,9511	0,9507	0,9542
U -Test	7	0,01	0,9657	0,9637	0,9618	0,9602	0,9587	0,9558	0,9544	0,9552
C_1-Test	7	0,01	0,9815	0,9681	0,9658	0,9635	0,9613	0,9562	0,9517	0,9403
U -Test	7	0,05	0,9508	0,9495	0,9484	0,9476	0,9470	0,9466	0,9476	0,9528
C_1-Test	7	0,05	0,9611	0,9594	0,9579	0,9566	0,9555	0,9536	0,9530	0,9542
I	5	0,01	0,9833	0,9679	0,9541	0,9422	0,9323	0,9185	0,9212	0,9620
	5	0,01	0,9833	0,9679	0,9541	0,9422	0,9323	0,9185	0,9212	0,9620
	6	0,01	0,9837	0,9695	0,9573	0,9475	0,9403	0,9347	0,9458	0,9858
	6	0,01	0,9847	0,9711	0,9593	0,9496	0,9422	0,9355	0,9452	0,9847
	7	0,01	0,9826	0,9684	0,9573	0,9496	0,9451	0,9476	0,9650	0,9960
	7	0,01	0,9906	0,9722	0,9617	0,9538	0,9486	0,9481	0,9629	0,9947

Beim Vergleich der A.R.E. (0,955) und der F.R.E.'s des U-Tests zum t-Test fällt eine nur geringe Abweichung der Effizienzen voneinander auf; das gilt nicht in gleichem Maße für den C_1-Test, dessen A.R.E. zum t-Test gleich 1 ist für Normalalternativen.

Insgesamt gesehen haben beide Rangtests hohe F.R.E.'s zum t-Test, insbesondere für $n = m = 3$. Diese sind im wesentlichen unabhängig von α, d und $m = n = 4, 5, 6, 7$; bis $d = 1,5$ nehmen sie geringfügig ab, um dann für $d = 2,0$ und $d = 3,0$ teilweise wieder leicht anzusteigen.

Die F.R.E.'s des U-Tests und die des C_1-Tests stimmen für $n = m = 3,4$ und $n = m = 5$ ($\alpha = 0,01$) überein. Sie unterscheiden sich sonst nur um weniger als eine Einheit in der 2. Stelle nach dem Komma. Dabei sind die F.R.E.'s des C_1-Tests bis auf Werte für d "am Rande" ($d = 2,0$ und $d = 3,0$) stets größer als die des U-Tests. Die Differenz zwischen den F.R.E.'s des C_1-Tests und des U-Tests scheint allerdings bei festem α für $d = 0,2$ (also in der Nähe der Nullhypothese) mit wachsendem n zuzunehmen, was wegen der größeren (lokalen) A.R.E. des C_1-Tests nicht überraschend ist.

5.4.2 F.R.E. für $m \neq n$ und einseitige Alternativen

Die F.R.E.'s in Tabelle C sind unter der Bedingung $m = n$ berechnet worden, und zwar nicht nur für den Stichprobenumfang der beiden Rangtests, sondern auch für den des t-Tests bei der linearen Interpolation zwischen $\bar{n}_2$ und $\bar{n}_2 + 1$. Diese Einschränkung wird in Tabelle D aufgegeben. In diesem Fall geht in Definition 16 für n_1 der Ausdruck $m+n$ ein. Dabei ist zu beachten, daß nun die lineare Interpolation zwischen den Stichprobenumfängen des t-Tests nicht mehr eindeutig ist.

So kann z.B. die Berechnung der F.R.E.'s der beiden Rangtests für (7,5) durch lineare Interpolation zwischen den Stichprobenumfängen des t-Tests (7,5) und (6,5) oder (7,5) und (7,4) erfolgen. In den Spalten 2 und 3 der Tabelle D sind die diesbezüglichen Stichprobenumfänge (m_1, n_1) und (m_2, n_2) für den t-Test angegeben.

Die F.R.E.'s des U-Tests und des C_1-Tests können - wie Tabelle D zeigt - im Vergleich zu den Werten in C deutlich kleiner werden und stark von den asymptotischen Ergebnissen abweichen, wenn α und die Differenz der beiden Stichprobenumfänge extrem groß sind. Dabei zeigt das Beispiel (7,2), welche unterschiedlichen F.R.E.'s sich durch die lineare Interpolation zwischen den verschiedenen Stichprobenumfängen bei dem t-Test ergeben.[1] Die für (7,5) und (7,6) tabellierten Werte unterscheiden sich nicht nennenswert von denen in Tabelle C ; beim U-Test ist wieder eine gute Übereinstimmung mit der A.R.E. festzustellen. Das gleiche gilt auch für $3 \leq m, n \leq 6$, wenn die Differenz zwischen m und n nicht größer als 2 ist. Auf die Tabellierung der diesbezüglichen F.R.E.'s ist hier verzichtet worden.

---

1) Da die Güte des t-Tests für (6,2), $\alpha = 0{,}25$ und alle d größer ist als die des U-Tests und des C_1-Tests für (7,2), muß also hier zwischen (6,2) und (5,2) interpoliert werden.

Tabelle D

F.R.E. zum t-Test bei Normalalternativen für ungleiche Stichprobenumfänge (einseitige Alternativen)

	(m,n)	$(m_1,n_1);(m_2,n_2)$	α	d=0,2	0,4	0,6	0,8	1,0	1,5	2,0	3,0
U -Test	(7,1)	(6,1) ; (5,1)	0,25	0,8105	0,8089	0,8073	0,8061	0,8040	0,7997	0,7953	0,7860
C_1-Test				0,8105	0,8089	0,8073	0,8061	0,8040	0,7997	0,7953	0,7860
U -Test	(7,2)	(7,2) ; (7,1)	0,25	0,9801	0,9795	0,9792	0,9791	0,9791	0,9801	0,9821	0,9881
C_1-Test				0,9822	0,9818	0,9815	0,9815	0,9816	0,9828	0,9850	0,9913
U -Test	(7,2)	(6,2) ; (5,2)	0,25	0,7922	0,7879	0,7831	0,7778	0,7720	0,7542	0,7298	0,6390
C_1-Test				0,8119	0,8081	0,8041	0,8000	0,7956	0,7832	0,7681	0,7225
U -Test	(7,5)	(7,5) ; (7,4)	0,01	0,9750	0,9731	0,9713	0,9695	0,9677	0,9633	0,9590	0,9489
C_1-Test				0,9762	0,9744	0,9727	0,9709	0,9691	0,9648	0,9605	0,9507
U -Test	(7,5)	(7,5) ; (6,5)	0,01	0,9581	0,9554	0,9528	0,9501	0,9473	0,9398	0,9312	0,9051
C_1-Test				0,9602	0,9577	0,9551	0,9524	0,9496	0,9423	0,9338	0,9084
U -Test	(7,5)	(7,5) ; (7,4)	0,05	0,9650	0,9639	0,9628	0,9618	0,9609	0,9587	0,9567	0,9516
C_1-Test				0,9721	0,9708	0,9695	0,9683	0,9672	0,9644	0,9616	0,9529
U -Test	(7,5)	(7,5) ; (6,5)	0,05	0,9373	0,9356	0,9338	0,9320	0,9300	0,9243	0,9169	0,8901
C_1-Test				0,9500	0,9476	0,9458	0,9435	0,9412	0,9347	0,9264	0,8931
U -Test	(7,6)	(7,6) ; (7,5)	0,01	0,9732	0,9714	0,9696	0,9679	0,9661	0,9616	0,9569	0,9437
C_1-Test				0,9758	0,9739	0,9721	0,9703	0,9684	0,9635	0,9580	0,9408
U -Test	(7,6)	(7,6) ; (6,6)	0,01	0,9657	0,9637	0,9616	0,9594	0,9572	0,9514	0,9448	0,9241
C_1-Test				0,9691	0,9669	0,9648	0,9625	0,9601	0,9538	0,9461	0,9208
U -Test	(7,6)	(7,6) ; (7,5)	0,05	0,9608	0,9598	0,9588	0,9578	0,9622	0,9545	0,9520	0,9444
C_1-Test				0,9668	0,9653	0,9637	0,9622	0,9606	0,9566	0,9520	0,9361
U -Test	(7,6)	(7,6) ; (6,6)	0,05	0,9486	0,9473	0,9460	0,9447	0,9504	0,9396	0,9348	0,9181
C_1-Test				0,9565	0,9545	0,9525	0,9505	0,9483	0,9424	0,9349	0,9058

5.4.3 F.R.E. für zweiseitige Alternativen

Bislang sind meines Wissens in der Literatur noch keine F.R.E.'s des U-Tests und des C_1-Tests für zweiseitige Alternativen angegeben worden. Der Vollständigkeit wegen und um auch die so gewonnenen Ergebnisse mit denen für einseitige Alternativen zu vergleichen, sind in Tabelle E F.R.E.'s für zweiseitige Alternativen $d = | \theta_Y - \theta_X |$ zusammengestellt, und zwar für ausgewählte $(m, n) - \alpha$ - Kombinationen aus den Tabellen C und D .

Der Vergleich mit den entsprechenden Werten in C und D bekräftigt die Vermutung, daß sich durch eine einseitige oder zweiseitige Fragestellung kein nennenswerter Unterschied für die F.R.E.'s ergibt.

Tabelle E

F.R.E.'s zum t-Test bei Normalalternativen für gleiche und ungleiche Stichprobenumfänge (zweiseitige Alternativen)

	(m,n)	$(m_1,n_1);(m_2,n_2)$	α	d=0,2	0,4	0,6	0,8	1,0	1,5	2,0	3,0
U -Test	(3,3)	(3,3) ; (2,2)	0,1	0,9776	0,9771	0,9764	0,9756	0,9747	0,9729	0,9724	0,9773
C_1-Test				0,9776	0,9771	0,9764	0,9756	0,9747	0,9729	0,9724	0,9773
U -Test	(7,7)	(7,7) ; (6,6)	0,01	0,9691	0,9675	0,9655	0,9634	0,9613	0,9567	0,9533	0,9492
C_1-Test				0,9709	0,9692	0,9670	0,9647	0,9624	0,9570	0,9521	0,9422
U -Test	(7,1)	(6,1) ; (5,1)	0,25	0,8571	0,8564	0,8554	0,8541	0,8524	0,8471	0,8406	0,8257
C_1-Test				0,8571	0,8564	0,8554	0,8541	0,8524	0,8471	0,8406	0,8257
U -Test	(7,5)	(7,5) ; (7,4)	0,01	0,9756	0,9745	0,9729	0,9712	0,9693	0,9647	0,9602	0,9504
C_1-Test				0,9756	0,9745	0,9730	0,9712	0,9654	0,9648	0,9603	0,9505
U -Test	(7,5)	(7,5) ; (6,5)	0,01	0,9606	0,9590	0,9568	0,9542	0,9515	0,9443	0,9362	0,9138
C_1-Test				0,9606	0,9591	0,9569	0,9543	0,9516	0,9444	0,9363	0,9141

5.4.4 F.R.E. über normale Interpolation

Die bislang zusammengestellten F.R.E.'s in den Tabellen C , D und E sind durch lineare Interpolation zwischen den Stichprobenumfängen für den t-Test berechnet worden. Es stellt sich jedoch die Frage, inwieweit gerade eine solche Interpolation gerechtfertigt ist, da bekanntlich die Gütefunktion des t-Tests für Normalalternativen generell keine konvexe Funktion in n ist. Der geringe Rechenaufwand spricht jedoch für eine lineare Interpolation.

Im folgenden sollen nun als Vergleich zum linearen Interpolationsverfahren F.R.E.'s nach Definition 16 dadurch erklärt werden, daß $\tilde{n}_2 \in \mathbb{R}^+$ über eine nicht-lineare Interpolation zwischen $\bar{n}_2$ und $\bar{n}_2 + 1$ bestimmt wird. Damit kann der Einfluß solcher "willkürlichen" Interpolationsverfahren auf die Definition der F.R.E. gerade für kleine Stichprobenumfänge untersucht werden. Die Art der Interpolation fällt bei großen Stichprobenumfängen nicht ins Gewicht.

Die Interpolation wird nun über $\Phi^{-1}\left[\beta_t(\tilde{n}_2, \alpha, \theta)\right]$ durchgeführt, wobei Φ die standardisierte Normalverteilung ist (im folgenden kurz normale Interpolation genannt) und β_t die Gütefunktion des t-Tests bezeichnet. Dieses Interpolationsverfahren ist deshalb naheliegend, weil die Güte-Kurve des t-Tests annähernd die Gestalt der Normalkurve hat.

Tabelle F

Vergleich der F.R.E.'s des U-Tests zum t-Test bei normaler (1.Zeile) und linearer (2.Zeile) Interpolation

(m,n)	$(m_1,n_1);(m_2,n_2)$	α	d=0,2	0,4	0,6	0,8	1,0	1,5	2,0	3,0
(3,3)	(3,3) ; (2,2)	0,1	0,9794	0,9782	0,9769	0,9756	0,9742	0,9703	0,9660	0,9562
			0,9787	0,9771	0,9756	0,9743	0,9731	0,9715	0,9719	0,9791
(4,4)	(4,4) ; (3,3)	0,1	0,9590	0,9573	0,9555	0,9536	0,9516	0,9461	0,9397	0,9233
			0,9582	0,9561	0,9542	0,9526	0,9514	0,9499	0,9508	0,9593
(5,5)	(5,5) ; (4,4)	0,05	0,9605	0,9590	0,9575	0,9558	0,9540	0,9491	0,9434	0,9283
			0,9597	0,9576	0,9559	0,9543	0,9531	0,9513	0,9515	0,9573
(6,6)	(6,6) ; (5,5)	0,05	0,9514	0,9460	0,9462	0,9456	0,9450	0,9408	0,9359	0,9228
			0,9506	0,9447	0,9449	0,9446	0,9440	0,9440	0,9451	0,9516
(7,7)	(7,7) ; (6,6)	0,01	0,9666	0,9651	0,9637	0,9621	0,9604	0,9557	0,9501	0,9358
			0,9657	0,9637	0,9618	0,9602	0,9587	0,9558	0,9544	0,9552
(7,7)	(7,7) ; (6,6)	0,05	0,9514	0,9504	0,9493	0,9482	0,9469	0,9432	0,9386	0,9263
			0,9508	0,9495	0,9484	0,9476	0,9470	0,9466	0,9476	0,9528
(7,1)	(7,1) ; (6,1)	0,25	0,8572	0,8565	0,8554	0,8525	0,8502	0,8471	0,8403	0,8239
			0,8571	0,8564	0,8554	0,8541	0,8524	0,8471	0,8406	0,8257

In Tabelle F sind die (m, n) - α - Kombinationen aus Tabelle C aufgenommen, für die (bei linearer Interpolation) ab d = 2,0 wachsende F.R.E.'s zu verzeichnen sind und dazu das Beispiel (7,1) mit den nach Tabelle E sehr niedrigen F.R.E.'s. Da hier nur ein Vergleich zwischen normaler und linearer Interpolation gezogen werden soll, sind die F.R.E.'s über normale Interpolation nur für den U-Test berechnet worden.

Die durch die beiden Arten der Interpolation bedingten F.R.E.-Unterschiede sind sehr gering bis d = 1,5 , nehmen dann zu und sind für d = 3,0 nicht mehr unerheblich. Auffällig ist, daß die F.R.E.'s über lineare Interpolation bis d = 1,0 kleiner sind, dann aber größer werden als die F.R.E.'s über normale Interpolation, die mit wachsendem d schnell abnehmen. Das widerspricht der geäußerten Vermutung, daß im allgemeinen die F.R.E.'s dieser Rangtests zum t-Test mit wachsendem d bei festen n und α wieder ansteigen. Die Art der Interpolation spielt also bei diesem Problem eine entscheidende Rolle.

5.4.5 F.R.E. des U-Tests für große m = n

Die bislang tabellierten F.R.E.'s bezogen sich auf kleine Stichprobenumfänge $n \leq 7$. Es erhebt sich die Frage, inwieweit die gute Übereinstimmung der F.R.E.'s mit der A.R.E. auch für große Stichprobenumfänge zutrifft. Im folgenden soll diese Frage für den U-Test im Vergleich zum t-Test für einseitige Normalalternativen $d > o$ beantwortet werden. Dabei werden Alternativen in der Nähe der Nullhypothese betrachtet, um einen Vergleich mit der entsprechenden A.R.E. (0,955) anstellen zu können, die gerade für Alternativen $d \to o$ definiert

ist. Während die F.R.E.'s bei den Stichprobenumfängen $n \leq 7$ über die exakten Gütewerte der betrachteten Tests berechnet worden sind, ist das für große $m = n$ wohl kaum möglich, denn bereits für $m = n = 10$ ergeben sich $\binom{10+10}{10} = 184756$ verschiedene Kombinationen der Ränge y_j (oder x_i) . So sollen nun für große n die Gütewerte des U-Tests approximativ über die Normalverteilung bestimmt werden. Damit wird die Tatsache ausgenutzt, daß die U-Statistik sowohl unter der Nullhypothese als auch unter der Alternativhypothese asymptotisch normalverteilt ist.

Statt der bislang verwendeten Statistik $U = \sum_{j=1}^{n} t_j$, wobei die t_j die geordneten Ränge der y_j bedeuten ($j = 1, \ldots, n$), wird im folgenden die Statistik

$$\bar{U} = \sum_{i=1}^{m} \sum_{j=1}^{n} Z_{ij} \qquad \text{benutzt mit}$$

$$Z_{ij} = \begin{cases} 1 & \text{für } X_i < Y_j \\ 0 & \text{für } X_i > Y_j \end{cases} \qquad i=1,\ldots,m \; ; \; j=1,\ldots,n \; .$$

Bekanntlich ist $\bar{U}$ äquivalent zu U , denn es gilt:

$$\bar{U} = U - \frac{n(n+1)}{2} \; .^{1)}$$

Die Berechnung der Güte des $\bar{U}$-Tests ergibt sich nun wie folgt:
Es ist unter $H_o : d = 0$ [2)]

(3) $$E_o(\bar{U}) = \frac{1}{2} mn$$

(4) $$Var_o(\bar{U}) = \frac{mn(m+n+1)}{12} \; .$$

---

1) Lehmann, [36], S. 253
2) Lehmann, [36], S. 254

Da für die hier betrachteten einseitigen Alternativen der kritische Bereich von $\bar{U}$ durch

$$P(\bar{U} > \bar{u}_\alpha) = \alpha$$

bestimmt wird und $\bar{U}$ unter H_o asymptotisch normalverteilt ist, kann $\bar{u}_\alpha$ approximativ dargestellt werden durch:

$$(5) \qquad \bar{u}_\alpha = E_o(\bar{U}) + z_\alpha \sqrt{Var_o(\bar{U})} \; ;$$

z_α ist dabei bestimmt durch $P(Z > z_\alpha) = \alpha$, wobei Z nach $N(0,1)$ verteilt ist.

Weiterhin gilt unter $H_1 : d > 0$ [1)]

$$(6) \qquad E_d(\bar{U}) = mnp_1$$

$$(7) \qquad Var_d(\bar{U}) = mn\left[p_1 + (m+n-2)\,p_2 - (m+n-1)\,p_1^2\right]$$

mit

$$p_1 = \int_0^1 F_1 dF_2 = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{d}{\sqrt{2}}} e^{-\frac{z^2}{2}}\, dz \qquad \text{und}$$

$$p_2 = \int_0^1 F_1^2 dF_2 = \int_{-\infty}^{\frac{d}{\sqrt{2}}} \int_{-\infty}^{\frac{d}{\sqrt{2}}} \frac{1}{2\pi(1-r^2)}\, e^{-\frac{1}{2(1-r^2)}\left[z_1^2 - 2rz_1z_2 + z_2^2\right]} dz_1 dz_2,$$

wobei $r = \frac{1}{2}$ ist.

---

1) Lehmann, [36], S. 253

Zu vorgegebenen α und d werden also zunächst p_1 und p_2 aus den entsprechenden Tabellen der eindimensionalen bzw. zweidimensionalen Normalverteilung bestimmt und dann $E_d(\bar{U})$ und $Var_d(\bar{U})$ berechnet.

Da $\bar{U}$ auch unter H_1 asymptotisch normalverteilt ist, kann also die Güte des $\bar{U}$-Tests [1] wegen (5) approximativ bestimmt werden durch:

$$P\left(Z > \frac{\bar{u}_\alpha - E_d(\bar{U})}{Var_d(\bar{U})}\right) =$$

$$(8) \qquad P\left(Z > \frac{E_o(\bar{U}) + z_\alpha \sqrt{Var_o(\bar{U})} - E_d(\bar{U})}{\sqrt{Var_d(\bar{U})}}\right) ,$$

wobei Z nach $N(0,1)$ verteilt ist. Die damit zu vergleichenden Gütewerte des t-Tests zur Berechnung der F.R.E.'s - hier wieder über lineare Interpolation - sind mit Hilfe einer Approximationsformel von Hodges-Lehmann [2] ermittelt worden.

---

1) und damit auch die des U-Tests

2) Hodges/Lehmann, [29], S. 1630

Tabelle G

F.R.E. des U-Tests zum t-Test bei einseitigen Normalalternativen für große Stichprobenumfänge (m=n)

m=n	α	d=0,2	0,4	0,6
20	0,01	0,9604	0,9244	0,9093
21	"	0,9591	0,9246	0,9108
22	"	0,9580	0,9248	0,9123
23	"	0,9570	0,9250	0,9137
24	"	0,9561	0,9253	0,9151
25	"	0,9553	0,9257	0,9165
30	"	0,9524	0,9275	0,9230
35	"	0,9507	0,9295	0,9287
40	"	0,9496	0,9315	0,9336
45	"	0,9488	0,9333	0,9379
50	"	0,9483	0,9350	0,9417
60	"	0,9478	0,9380	0,9481
70	"	0,9476	0,9406	0,9534
80	"	0,9476	0,9428	0,9579
20	0,05	0,9392	0,9213	0,9230
21	"	0,9393	0,9224	0,9251
22	"	0,9394	0,9235	0,9272
23	"	0,9396	0,9245	0,9292
24	"	0,9397	0,9255	0,9310
25	"	0,9398	0,9265	0,9328
30	"	0,9406	0,9307	0,9405
35	"	0,9413	0,9343	0,9467
40	"	0,9420	0,9372	0,9518
45	"	0,9427	0,9397	0,9562
50	"	0,9433	0,9419	0,9599
60	"	0,9443	0,9438	0,9661
70	"	0,9452	0,9455	0,9710
80	"	0,9459	0,9484	0,9749

Die tabellierten F.R.E.'s zeigen zumindest für $d = 0{,}2$ eine gute Übereinstimmung mit der A.R.E. und bestätigen damit auch für große Stichprobenumfänge die hohe Effizienz des U-Tests zum t-Test. Bemerkenswert ist noch ein Vergleich mit den Ergebnissen von H. Witting [1], der eine Verallgemeinerung der Pitmanschen A.R.E. über die Edgeworth-Approximation einführt.

Die Wittingschen A.R.E.'s werden für verschiedene $(m, n) - \alpha$ - Kombinationen berechnet, beziehen sich aber wie die Pitmansche A.R.E. auf Alternativen $d \to 0$ [2], so daß also die A.R.E.'s nach Witting für wachsendes n eine bessere Übereinstimmung mit denen von Pitman erwarten lassen als die hier tabellierten F.R.E.'s selbst für $d = 0{,}2$. Das bestätigt die folgende Tabelle, in der die Wittingschen A.R.E.'s (3. Spalte) mit den entsprechenden F.R.E.'s aus Tabelle G zusammengestellt sind:

m=n		A.R.E.	F.R.E.
20	0,01	0,9566	0,9604
40	0,01	0,9558	0,9496
20	0,05	0,9548	0,9392
40	0,05	0,9521	0,9420

5.4.6 Abschließende Bemerkungen

Mit den Untersuchungen dieses Kapitels ist für den U-Test und C_1-Test der Nachweis erbracht, daß bei kleinen Stichprobenumfängen die finiten

1) Witting, [70], S. 405

2) für $m,n \to \infty$ stimmt das Konzept von Witting mit dem von Pitman überein.

relativen Effizienzen zum t-Test kaum von der asymptotischen Effizienz abweichen, abgesehen von extremen $(m,n) - \alpha$ - Kombinationen. Diese gute Übereinstimmung der F.R.E.'s mit der A.R.E. wird im vorangegangenen Abschnitt für den U-Test auch bei großen Stichprobenumfängen in der Nähe der Nullhypothese bestätigt.

Die Berechnung der F.R.E.'s ist recht mühsam, da zur Bestimmung eines jeden Wertes 3 Variable (α, n, θ) berücksichtigt werden müssen. Die A.R.E. als eine einzige Zahl hat dagegen allgemeinen Charakter. So kann der Unterschied zwischen F.R.E. und A.R.E. kurz folgendermaßen gekennzeichnet werden:

Die F.R.E. ist realistisch, aber wenig allgemein, die A.R.E. allgemein, aber wenig realistisch. - Völlig offen ist bislang noch das Problem, wie hoch die F.R.E.'s des U-Tests, des C_1-Tests oder eines anderen Rangtests auf Lage im Vergleich zum t-Test bei Annahme nicht-normalverteilter Grundgesamtheiten sind und damit die Frage, inwieweit solche F.R.E.'s mit den entsprechenden, in Tabelle B aufgeführten A.R.E.'s übereinstimmen. Einige Ansätze zur Berechnung der Gütewerte des U-Tests für eine Rechteck- und Exponentialverteilung[1] sowie für eine Doppelexponentialverteilung[2] (auch für andere Rangstatistiken) liegen bereits vor.

Noch wichtiger als die Lösung dieses umfangreichen Problems scheint mir die Aufgabe zu sein, auch für Rangtests auf Variabilität im Vergleich zum F-Test finite relative Effizienzen bezüglich Normalalternativen zu berechnen. Dazu gibt es meines Wissens bislang - abgesehen von einer einzigen Veröffentlichung von J. Klotz[3] - keinerlei Untersuchungen. Klotz zeigt darin, daß die F.R.E.'s des nach ihm

---

1) Haynam/Govindarajulu, [26], S. 945

2) Ramsey, [51], S. 149

3) Klotz, [32], S. 498

benannten K-Tests zum F-Test für $m = n = 5$, $\alpha = \frac{16}{\binom{10}{5}} \simeq 0{,}064$ und $\frac{\sigma}{\tau} = 1{,}5;\ 2{,}0;\ 3{,}0;\ 4{,}0$ von 0,803 bis auf 0,640 fallen, also viel niedriger sind als die entsprechenden F.R.E.'s des U-Tests oder C_1-Tests zum t-Test. Es bliebe die These zu prüfen, daß dies auch für weitere (m,n) - α - Kombinationen und andere Rangtests auf Variabilität zutrifft.

Literaturverzeichnis

Abkürzungen: AMS : Annals of Mathematical Statistics
JASA: Journal of the American Statistical Association

[1] Bahadur, R.R. — Stochastic comparison of tests
AMS 31 (1960), S.276

[2] Basu, A.B.
Woodworth, G. — A note on nonparametric tests for scale
AMS 38 (1967), S.274

[3] Behnen, K. — Asymptotic optimality and ARE of certain rank-order tests under contiguity
AMS 42 (1971), S.225

[4] Bell, C.B.
Doksum, K.A. — Some new distribution-free statistics
AMS 36 (1965), S.203

[5] Bell, C.B.
Doksum, K.A. — Optimal one sample distribution-free tests and their two-sample extensions
AMS 37 (1966), S.120

[6] Bell, C.B.
Moser, J.M.
Thompson, R. — Goodness criteria for two-sample distribution-free tests
AMS 37 (1966), S.133

[7] Beran, R.L. — On distribution-free statistical inference with upper und lower probabilities
AMS 42 (1971), S.157

[8] Blyth, C.R. — Note of relative efficiency of tests
AMS 29 (1958), S.898

[9] Bradley, J.V. — Distribution-free statistical tests
Prentice-Hall, New Yersey 1968

[10]	Capon, J.	Asymptotic efficiency of certain locally most powerful rank tests AMS 32 (1961), S.88
[11]	Chernoff, H. Savage, I.R.	Asymptotic normality and efficiency of certain non parametric test statistics AMS 29 (1958), S.972
[12]	Cramér, H.	Mathematical methods of statistics Princeton 1946
[13]	Dixon, W.J.	Power functions of the sign test and power efficiency for normal alternatives AMS 24 (1953), S.467
[14]	Dixon, W.J.	Power under normality of several nonparametric tests AMS 25 (1954), S.610
[15]	Dwass, M.	On the asymptotic normality of some statistics used in non parametric tests AMS 26 (1955), S.334
[16]	v.Eeden, C.	Note on the consistency of some distribution-free tests for dispersion JASA (1964), S.105
[17]	Fisz, M.	Wahrscheinlichkeitsrechnung und Mathematische Statistik VEB Deutscher Verlag der Wissenschaften, Berlin 1966
[18]	Fraser, D.A.S.	Non-parametric methods in statistics J. Wiley, New York 1957
[19]	Gastwirth, J.L. Wolff, S.S.	An elementary method for obtaining lower bounds on the asymptotic power of rank tests AMS 39 (1968), S.2128

[20]	Gastwirth, J.L.	On robust procedures JASA 61 (1966), S.929
[21]	Gibbons, J.D.	On the power of two-sample rank tests on the equality of two distribution functions Journal of the Royal Statistical Society, Series B, 26, S.293
[22]	Gumbel, E.J.	Statistics of extremes Columbia University Press, New York 1960
[23]	Hajek, J.	Asymptotic efficiency of a certain sequence of tests (russisch) Czechoslovak Mathematical Journal 6 (81), S.26
[24]	Hajek, J. Sidak, Z.	Theory of rank tests Academic press, New York, 1967
[25]	Hajek, J.	Non parametric Statistics Holden-Day, San Francisco, 1969
[26]	Haynam, G.E. Govindarajulu, Z.	Exact power of the U-Test for exponential and rectangular alternatives AMS 37 (1966), S.945
[27]	Hodges, J.L. Lehmann, E.L.	The efficiency of some non-parametric competitors of the t-Test AMS 27 (1956), S.324
[28]	Hodges, J.L. Lehmann, E.L.	Comparison of the normal scores and Wilcoxon test Proceedings IVth Berkeley Symposium I, S.307
[29]	Hodges, J.L. Lehmann, E.L.	A compact table for power of the t-Test AMS 39 (1968), S.1629

[30] Kendall, M.G.
Rank correlation methods
Griffin & Co., London 1962

[31] Kendall, M.G.
Sundrum, R.M.
Distribution-free Methods and order properties
Revue de l'institut international de statistique 21 (1953), S.124

[32] Klotz, J.
Nonparametric tests for scale
AMS 33 (1962), S.498

[33] Klotz, J.
On the normal scores two-sample rank test
JASA 59 (1964), S.652

[34] Lauscher, N.F.
Steffens, F.E.
de Lange, E.M.
Exact critical values for Mood's distribution-free tests statistic for dispersion and its normal approximation
Technometrics 10 (1968), S.497

[35] Lehmann, E.L.
The power of rank tests
AMS 24 (1953), S.23

[36] Lehmann, E.L.
Testing statistical hypotheses
J. Wiley, New York 1959

[37] Lienert, G.A.
Verteilungsfreie Methoden in der Biostatistik
Meisenheim am Glan 1962

[38] Lindgren, B.W.
Statistical Theory
The Macmillan Company, New York 1965

[39] Mann, H.B.
Whitney, D.R.
On a test of whether one of two random variables is stochastically larger than the other
AMS 18 (1947), S.50

[40]	Milton, R.C.	Rank order probabilities, two-sample normal shift alternatives J. Wiley, New York 1970
[41]	Mood, A.M.	On the asymptotic efficiency of certain nonparametric two-sample-tests AMS 25 (1954), S.514
[42]	Mood, A.M. Graybill, F.A.	Introduction of the theory of statistics Mc Graw-Hill Book Company, New York 1963
[43]	Moses, L.E.	Rank tests of dispersion AMS 34 (1963), S.973
[44]	Noether, G.E.	On an theorem of Pitman AMS 26 (1955), S.64
[45]	Noether, G.E.	The efficiency of some distribution free tests Statistica Neerlandica 12 (1958), S.63
[46]	Pfanzagl, J.	Über lokal optimale Rangtests Metrika 3 (1960), S.143
[47]	Pfanzagl, J.	Allgemeine Methodenlehre der Statistik II Walter de Gruyter & Co., Berlin 1966
[48]	Pfanzagl, J.	Theory of Measurement Physica-Verlag, Würzburg 1968
[49]	Puri, M.L. Sen, P.K.	Nonparametric methods in multivariate Analysis J. Wiley, New York 1971
[50]	Raghavachari, M.	On the efficiency of the normal scores test relative to the F-Test AMS 36 (1965), S.1306

[51]	Ramsey, F.L.	Small sample power functions for nonparametric tests of location in the double exponential case JASA 66 (1971), S.149
[52]	Rao, C.R.	Linear statistical inference and its applications J. Wiley, New York 1965
[53]	Rytz, C.	Ausgewählte parameterfreie Prüfverfahren im 2- und k-Stichproben-Fall Metrika 12 (1967), S.189
[54]	Rytz, C.	Ausgewählte parameterfreie Prüfverfahren im 2- und k-Stichproben-Fall Metrika 13 (1968), S.17
[55]	Savage, I.R.	Contributions to the theory of rank order statistics - two-sample case AMS 27 (1956), S.590
[56]	Savage, I.R.	Bibliography of nonparametric statistics Harvard University Press, 1962
[57]	Schmetterer, L.	Über nichtparametrische Methoden in der Mathematischen Statistik Jahresbericht der Deutschen Mathematiker Vereinigung 61 (1959), S.104
[58]	Schmetterer, L.	Mathematische Statistik Springer-Verlag, Wien 1966
[59]	Siegel, S.	Nonparametric statistics for the behavioral sciences Mc Graw-Hill, New York 1956
[60]	Srivastava, A.B.L.	Effect of non-normality on the power function of the t-Test Biometrika 45 (1958), S.421

[61]	Sukhatme, B.V.	On certain two-sample non-parametric tests for variances AMS 28 (1957), S.188
[62]	Tsutakawa, R.K.	An example of large discrepancy between measures of asymptotic efficiency of tests AMS 39 (1968), S.179
[63]	Uzawa, H.	Locally most powerful rank tests for two-sample problems AMS 31 (1960), S.685
[64]	Verdooren, L.R.	Extended tables of critical values of Wilcoxon's test statistic Biometrika 50 (1963), S.177
[65]	v.d.Waerden, B.L.	Algebra I Springer-Verlag, Berlin 1955
[66]	v.d.Waerden, B.L.	Order tests for the two-sample problem and their power Proceedings Koninkl. Nederl. Akad.Wetenschapen, 56 (1953), S.303 und S.311
[67]	Walsh, J.E.	Handbook of nonparametric Statistics Van Nostran, Princeton 1962
[68]	Witting, H.	Mathematische Statistik B.G. Teubner, Stuttgart 1969
[69]	Witting, H. Nölle, G.	Angewandte Mathematische Statistik B.G. Teubner, Stuttgart 1970
[70]	Witting, H.	A generalized Pitman efficiency for non-parametric tests AMS 31 (1960), S.405

Lebenslauf

Am 20.7.1939	als Sohn des Bäckermeisters Wilhelm Büning und seiner Frau Maria Büning, geb. Benner, in Oberhausen geboren
Im März 1958	Reifeprüfung am Naturwissenschaftlichen Gymnasium in Oberhausen
SS 1958 - WS 1959/60	Studium der Mathematik, Philosophie und Leibeserziehung in Marburg
SS 1960 - SS 1961	Fortsetzung des Studiums in Freiburg
WS 1961/1962 - WS 1963/1964	Fortsetzung des Studiums in Münster
Im Mai 1964	1. Staatsexamen in Münster
Im Oktober 1964	Beginn der Referendarzeit in Berlin
Im Juni 1966	2. Staatsexamen (Assessorexamen)
Im Oktober 1968	Ernennung zum Studienrat
Im Juni 1969	Ernennung zum Akademischen Rat an der Freien Universität Berlin
Im April 1971	Ernennung zum Professor an der Freien Universität Berlin
Im Dezember 1973	Promotion zum Dr. rer. pol. mit dem Hauptfach Statistik an der Freien Universität Berlin